GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 27, 2004, 10:02:56 ; Search time 54 Seconds
(without alignments)
1104.028 Million cell updates/sec

Title: US-09-813-820-4
Perfect score: 1122
Sequence: 1 MRGSHHHHHHGSDDKVATIT.....GIEGTVKGELKVLKQDKDTK 211

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1122	100.0	211	2	AAW31553	Aaw31553 Collagen
2	1054.5	94.0	512	2	AAW31554	Aaw31554 Collagen
3	1048	93.4	1183	6	ABU79084	Abu79084 S. aureus
4	904	80.6	1185	2	AAR22675	Aar22675 Collagen
5	884	78.8	168	4	AAE11855	Aae11855 Staphyloc
6	830	74.0	159	2	AAW31552	Aaw31552 Collagen
7	491	43.8	731	7	ADC97372	Adc97372 E. faeciu
8	178	15.9	458	3	AAB37667	Aab37667 Collagen
9	178	15.9	458	7	ADD90648	Add90648 Enterococ
10	165.5	14.8	146	3	AAB37668	Aab37668 Collagen
11	154	13.7	168	4	AAE11856	Aae11856 Enterococ
12	141.5	12.6	316	7	ABM79017	Abm79017 Staphyloc
13	131.5	11.7	343	7	ABM79016	Abm79016 Staphyloc
14	125.5	11.2	345	2	AAW31555	Aaw31555 Firbonect
15	119	10.6	336	5	ABB53995	Abb53995 Lactococc
16	118	10.5	560	7	ABM79015	Abm79015 Staphyloc
17	116	10.3	473	2	AAW68400	Aaw68400 Clostridi
18	114.5	10.2	688	2	AAY00241	Aay00241 Enterococ
19	114.5	10.2	688	5	ABP43460	Abp43460 E faecali
20	114.5	10.2	688	6	ABU88488	Abu88488 E. faecal
21	114.5	10.2	688	6	ABU13739	Abu13739 Enterococ
22	114.5	10.2	2032	2	AAY00238	Aay00238 Enterococ
23	114.5	10.2	2032	2	AAY00240	Aay00240 Enterococ
24	114.5	10.2	2032	2	AAY00242	Aay00242 Enterococ
25	114.5	10.2	2032	5	ABP43459	Abp43459 E faecali
26	114.5	10.2	2032	5	ABP43457	Abp43457 E faecali
27	114.5	10.2	2032	5	ABP43461	Abp43461 E faecali
28	114.5	10.2	2032	6	ABU88489	Abu88489 E. faecal
29	114.5	10.2	2032	6	ABU88485	Abu88485 E. faecal
30	114.5	10.2	2032	6	ABU88487	Abu88487 E. faecal
31	114.5	10.2	2032	6	ABU13740	Abu13740 Enterococ
32	114.5	10.2	2032	6	ABU13736	Abu13736 Enterococ
33	114.5	10.2	2032	6	ABU13738	Abu13738 Enterococ
34	114	10.2	200	6	ABP58754	Abp58754 Recombina
35	113.5	10.1	184	2	AAY32917	Aay32917 E.coli op
36	113.5	10.1	184	3	AAB10309	Aab10309 Human KGF
37	113.5	10.1	184	4	AAB61673	Aab61673 pQE60-Cys
38	113.5	10.1	184	6	ABP54289	Abp54289 pQE60-Cys
39	113.5	10.1	184	6	ADA95466	Ada95466 Synthetic
40	113.5	10.1	184	7	ADD66139	Add66139 Human mat
41	113.5	10.1	1788	6	ABU18740	Abu18740 Protein e
42	113	10.1	303	5	ABG67279	Abg67279 Polythoa
43	111.5	9.9	2627	6	ABR39483	Abr39483 L. cuprin
44	111.5	9.9	3333	6	ABR39482	Abr39482 L. cuprin
45	110.5	9.8	1338	2	AAR41731	Aar41731 High mole

ALIGNMENTS

RESULT 1
AAW31553
ID AAW31553 standard; protein; 211 AA.
XX
AC AAW31553;
XX
DT 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 21-MAY-1998 (first entry)
XX
DE Collagen binding protein M31 epitope.
XX
KW Collagen binding protein; cna gene; sepsis; infection;
KW microbial surface component regonising adhesive matrix molecule; MSCRAMM;
KW adhesin; vaccine; immunisation; diagnosis; therapy; epitope M31.
XX
OS Staphylococcus aureus.
XX
FH Key Location/Qualifiers
FT Peptide 1. .12
FT /note= "vector pQE30-derived peptide"
FT Protein 13. .211
FT /note= "epitope M31"
XX
PN WO9743314-A2.
XX
PD 20-NOV-1997.
XX
PF 14-MAY-1997; 97WO-US008210.
XX
PR 16-MAY-1996; 96US-0017678P.
XX
PA (TEXA) UNIV TEXAS A & M SYSTEM.
PA (UABR-) UAB RES FOUND.
XX
PI Hoeoek M, Patti JM, House-Pompeo K, Sthanam N, Symersky J;
XX
DR WPI; 1998-008801/01.
DR N-PSDB; AAT93437.
XX
PT Antibody that interacts with collagen binding domain of Staphylococcal
PT cna gene product - useful to prevent bacterial sepsis in animal infected
PT with Staphylococcus aureus.
XX
PS Claim 31; Page 115-116; 143pp; English.
XX
CC This protein comprises Staphylococcus aureus collagen binding protein
CC (CBP) epitope M31, i.e. amino acids 61-343 of full-length CBP, plus a

CC vector-derived N-terminal peptide. Claimed 441, 849 and 1500 bp nucleic
CC acid sequences (see AAT93436-38) respectively encode CBP epitopes M17,
CC M31 and M55 (see AAW31552-54) that confer protection against S. aureus
CC infection. These nucleic acid sequences can be used in the recombinant
CC production of the CBP epitopes. The CBP protein and antigenic epitopes
CC are contemplated for use in the treatment of pathological infections,
CC especially to prevent bacterial adhesion to collagen. The claimed nucleic
CC acids as well as claimed anti-CBP antibodies will also be of use in
CC screening, diagnostic and therapeutic applications including active and
CC passive immunisation and methods for the prevention of bacterial
CC colonisation in an animal such as a human. The CBP epitopes are also
CC contemplated for use in the preparation of vaccines and as carrier
CC proteins in vaccine formulations, as well as in the formulation of
CC compositions for the prevention of S. aureus infection. (Updated on 25-
CC MAR-2003 to correct PI field.) (Updated on 27-AUG-2003 to correct OS
CC field.)
XX
SQ Sequence 211 AA;

Query Match 100.0%; Score 1122; DB 2; Length 211;
Best Local Similarity 100.0%; Pred. No. 9.2e-86;
Matches 211; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MRGSHHHHHHGSDDKVATITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRGSHHHHHHGSDDKVATITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFL 60

Qy         61 NINNEKSYVSKDITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NINNEKSYVSKDITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKI 120

Qy        121 TVDNTKNTIDVTIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TVDNTKNTIDVTIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSF 180

Qy        181 NHTVHNINANAGIEGTVKGELKVLKQDKDTK 211
              |||||||||||||||||||||||||||||||
Db        181 NHTVHNINANAGIEGTVKGELKVLKQDKDTK 211
```

RESULT 2
AAW31554
ID AAW31554 standard; protein; 512 AA.
XX
AC AAW31554;
XX
DT 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 21-MAY-1998 (first entry)
XX
DE Collagen binding protein M55 epitope.
XX
KW Collagen binding protein; cna gene; sepsis; infection;
KW microbial surface component regonising adhesive matrix molecule; MSCRAMM;
KW adhesin; vaccine; immunisation; diagnosis; therapy; epitope M55.
XX
OS Staphylococcus aureus.
XX
FH Key Location/Qualifiers
FT Peptide 1. .12
FT /note= "vector pQE30-derived peptide"
FT Protein 13. .512
FT /note= "epitope M17"
XX
PN WO9743314-A2.
XX
PD 20-NOV-1997.
XX
PF 14-MAY-1997; 97WO-US008210.
XX
PR 16-MAY-1996; 96US-0017678P.
XX
PA (TEXA) UNIV TEXAS A & M SYSTEM.
PA (UABR-) UAB RES FOUND.
XX
PI Hoeoek M, Patti JM, House-Pompeo K, Sthanam N, Symersky J;
XX
DR WPI; 1998-008801/01.
DR N-PSDB; AAT93438.
XX
PT Antibody that interacts with collagen binding domain of Staphylococcal
PT cna gene product - useful to prevent bacterial sepsis in animal infected
PT with Staphylococcus aureus.
XX
PS Claim 31; Page 117-119; 143pp; English.
XX
CC This protein comprises Staphylococcus aureus collagen binding protein
CC (CBP) epitope M55, i.e. amino acids 30-531 of full-length CBP, plus a
CC vector-derived N-terminal peptide. Claimed 441, 849 and 1500 bp nucleic
CC acid sequences (see AAT93436-38) respectively encode CBP epitopes M17,
CC M31 and M55 (see AAW31552-54) that confer protection against S. aureus
CC infection. These nucleic acid sequences can be used in the recombinant
CC production of the CBP epitopes. The CBP protein and antigenic epitopes
CC are contemplated for use in the treatment of pathological infections,
CC especially to prevent bacterial adhesion to collagen. The claimed nucleic
CC acids as well as claimed anti-CBP antibodies will also be of use in
CC screening, diagnostic and therapeutic applications including active and
CC passive immunisation and methods for the prevention of bacterial
CC colonisation in an animal such as a human. The CBP epitopes are also
CC contemplated for use in the preparation of vaccines and as carrier
CC proteins in vaccine formulations, as well as in the formulation of
CC compositions for the prevention of S. aureus infection. (Updated on 25-
CC MAR-2003 to correct PI field.) (Updated on 27-AUG-2003 to correct OS
CC field.)
XX
SQ Sequence 512 AA;

Query Match 94.0%; Score 1054.5; DB 2; Length 512;
Best Local Similarity 64.7%; Pred. No. 1.3e-79;
Matches 211; Conservative 0; Mismatches 0; Indels 115; Gaps 1;

```
Qy          1 MRGSHHHHHHG---------------------------------------------------- 11
              ||||||||||
Db          1 MRGSHHHHHHGSARDISSTNVTDLTVSPSKIEDGGKTTVKMTFDDKNGKIQNGDMIKVAW 60

Qy         12 ------------------------------------------------------------ 11

Db         61 PTSGTVKIEGYSKTVPLTVKGEQVGQAVITPDGATITFNDKVEKLSDVSGFAEFEVQGRN 120

Qy         12 ------SDDKVATITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFLNINNE 65
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LTQTNTSDDKVATITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFLNINNE 180

Qy         66 KSYVSKDITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNT 125
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KSYVSKDITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNT 240

Qy        126 KNTIDVTIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVH 185
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KNTIDVTIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVH 300

Qy        186 NINANAGIEGTVKGELKVLKQDKDTK 211
              ||||||||||||||||||||||||||
Db        301 NINANAGIEGTVKGELKVLKQDKDTK 326
```

RESULT 3
ABU79084
ID ABU79084 standard; protein; 1183 AA.
XX
AC ABU79084;
XX
DT 18-JUN-2003 (first entry)
XX
DE S. aureus collagen adhesin protein.
XX

KW Superantigen; SAg; staphylococcal enterotoxin; tumour; cancer; apoptosis;
KW gene therapy; mammalian cell receptor; cytostatic;
KW tumour associated lipid; anergy; T cell; antigen presenting cell; APC;
KW tumouricidal immunocyte; antitumour.
XX
OS Staphylococcus aureus.
XX
PN US2002177551-A1.
XX
PD 28-NOV-2002.
XX
PF 30-MAY-2001; 2001US-00870759.
XX
PR 31-MAY-2000; 2000US-0208128P.
XX
PA (TERM/) TERMAN D S.
XX
PI Terman DS;
XX
DR WPI; 2003-361759/34.
DR N-PSDB; ACA64715.
XX
PT A mammalian cell receptor, useful in the treatment of cancer by binding
PT to tumor associated lipids where the binding induces anergy or apoptosis
PT in T cells and antigen presenting cells.
XX
PS Disclosure; Page; 167pp; English.
XX
CC The invention relates to a mammalian cell receptor, useful in the
CC treatment of cancer, which binds to tumour associated lipids and induces
CC anergy or apoptosis in the T cells and antigen presenting cells (APCs).
CC Also included are a mammalian cell useful in the treatment of cancer
CC where the receptor which binds tumour associated lipids and induces
CC cellular inactivation or death is deleted or functionally deactivated,
CC producing (M1) a tumouricidal immunocyte population in vivo in a mammal
CC (by allowing tumour associated lipids to contact immunocytes in which
CC receptors for immunosuppressive fatty acids, ceramides, glycolipids,
CC sphingolipids, glycosphingolipids, phosphosphingolipids, gangliosides,
CC sialylated glycans, lipopeptides and proteoglycolipids are inactivated or
CC deleted), a construct useful in the treatment of cancer comprising a
CC superantigen (SAg) nucleotide inserted into a virus, a mammalian T cell
CC useful in the treatment of cancer (where an adaptor protein which
CC inhibits T cell activation by tumour associated antigens is deleted or
CC functionally deactivated), a composition useful in the treatment of
CC cancer (comprising a lipid raft conjugated to a superantigen), producing
CC (M2) a tumouricidal immunocyte population ex vivo in a mammal (by
CC allowing tumour associated lipids to contact immunocytes, in which
CC receptors for the lipids are inactivated or deleted to produce a
CC tumouricidal immunocyte population, and administering the tumouricidally
CC activated immunocytes to the host), producing (M3) a tumouricidal APC
CC population ex vivo in a mammal (by allowing a tumour associated lipid to
CC contact APCs, in which receptors for the tumour associated lipids are
CC inactivated or deleted to produce a tumouricidally activated population,
CC and administering APCs to the host), producing a tumouricidal T cell
CC population ex vivo in a mammal) by allowing a tumour associated lipids to
CC contact T cells, in which adaptor proteins, which inhibit T cell
CC activation by tumour associated antigens, are deleted or functionally
CC deactivated to produce a tumouricidal population of T cells, and
CC administering the tumouricidally activated T cells to the host, or
CC allowing a superantigen-lipid raft to contact T cells ex vivo, and
CC administering the tumouricidally activated T cells to the host), treating
CC (M5) cancer in a mammal (by administering a lipid binding molecule which
CC binds immunosuppressive tumour associated lipids in vivo), producing (M6)
CC a tumouricidal T cell population in vivo in a mammal (by allowing a
CC tumour associated antigen to contact immunocytes in which adaptor
CC proteins which inhibit T cell activation by tumour associated antigens
CC are deleted or functionally deactivated) and producing (M7) a
CC tumouricidal T cell population ex vivo in a mammal comprising allowing a
CC superantigen-lipid raft conjugate to contact immunocytes in vivo. The
CC receptors, methods and compositions are useful for treating cancers and
CC tumours. Bacterial superantigens are co-administered or administered as
CC fusion constructs with anti-tumour proteins or motifs. The present
CC sequence represents an anti-tumour protein which is co-administered with
CC or incorporated into a fusion construct with a superantigen. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format from the US patnet
CC office website at "seqdata.uspto.gov/sequence.html?DocID=20020177551"
XX
SQ Sequence 1183 AA;

Query Match 93.4%; Score 1048; DB 6; Length 1183;
Best Local Similarity 100.0%; Pred. No. 1.4e-78;
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 SDDKVATITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSK 71
Db 144 SDDKVATITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSK 203

Qy 72 DITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDV 131
Db 204 DITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDV 263

Qy 132 TIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINANA 191
Db 264 TIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINANA 323

Qy 192 GIEGTVKGELKVLKQDKDTK 211
Db 324 GIEGTVKGELKVLKQDKDTK 343

RESULT 4
AAR22675
ID AAR22675 standard; protein; 1185 AA.
XX
AC AAR22675;
XX
DT 25-MAR-2003 (revised)
DT 04-DEC-1992 (first entry)
XX
DE Collagen binding protein.
XX
KW CBP; collagen binding protein; mastitis; arthritis.
XX
OS Staphylococcus aureus.
XX
FH Key Location/Qualifiers
FT Peptide 1. .29
FT /label= sig_peptide
FT Region 30. .1185
FT /note= "claim 11; page 29-30"
FT Region 30. .534
FT /label= A
FT /note= "see CC"
FT Region 535. .721
FT /label= B1
FT /note= "see CC"
FT Region 722. .908
FT /label= B2
FT /note= "see CC"
FT Region 909. .1095
FT /label= B3
FT /note= "see CC"
FT Region 1096. .1159
FT /label= W
FT /note= "see CC"
FT Region 1160. .1179
FT /label= M
FT /note= "see CC"
FT Region 1180. .1185
FT /note= "charged C-terminal"
XX
PN WO9207002-A1.
XX
PD 30-APR-1992.
XX

PF 22-OCT-1991; 91WO-SE000707.
XX
PR 22-OCT-1990; 90SE-00003374.
PR 22-OCT-1991; 91WO-SE000707.
XX
PA (ALFA) ALFA LAVAL AGRIC INT AB.
XX
PI Guss BM, Hook M, Jonsson H, Lindberg KM, Patti J, Signaes LC;
PI Switalski LM;
XX
DR WPI; 1992-167099/20.
DR N-PSDB; AAQ24123.
XX
PT Hybrid DNA molecule encoding S.aureus collagen binding protein - protein
PT is expressed in E.coli and used for diagnosis e.g. of septic arthritis.
XX
PS Disclosure; Fig 2; 40pp; English.
XX
CC The amino acid sequence (encoded by the sequence assembled from partially
CC homologous p16 and cCOLR6A inserts) comprises a structure resembling a
CC signal sequence. Following this region, a region called A is found
CC followed by a repetitive stretch of 187 amino acids called B1, B2 and B3.
CC Directly following these regions there is a region called W which
CC consists of a repetitive, hydrophilic structure contg. several proline
CC residues. This region is thought to mediate the binding of the protein to
CC the cell wall. The amino acid sequence nearest to the C-terminal end
CC consists of a long stretch of hydrophobic residues followed by some
CC charged amino acids This region is called M. The CBP can be used for
CC immunisation pref. in combination with a fusion protein, e.g. for
CC vaccination of ruminants against mastitis caused by staphylococcal
CC infections. It can also be used to block infection in an open skin wound,
CC e.g. for blocking protein receptors or by immunisation. In the latter,
CC the host produces specific antibodies which block the adherence of the
CC bacterial strains to damaged tissue. This treatment can be used for
CC septic arthritis and tissue damage of e.g. skin, connective tissue, and
CC mucous membranes. Dosage for immunisation is 0.5-5 microg CBP/kg; for
CC topical admin. the protein is used at a concn. of 25-250 microg/ml.
CC (Updated on 25-MAR-2003 to correct PN field.) (Updated on 25-MAR-2003 to
CC correct PF field.) (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ Sequence 1185 AA;

Query Match 80.6%; Score 904; DB 2; Length 1185;
Best Local Similarity 89.6%; Pred. No. 1.5e-66;
Matches 180; Conservative 3; Mismatches 16; Indels 2; Gaps 2;

```
Qy         13 DDKVATITSGNKSTNVTVHKSEAGTSSVFYY-KTGDM-LPEDTTHVRWFLNINNEKSYVS 70
              |||||||||||||||||           ||   |:| :   ||||||||||||||||||||
Db        145 DDKVATITSGNKSTNVIGWIKVKREPVVFLINKSGKICYQEDTTHVRWFLNINNEKSYVS 204

Qy         71 KDITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTID 130
              |||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        205 KDITIKDEIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTID 264

Qy        131 VTIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINAN 190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        265 VTIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINAN 324

Qy        191 AGIEGTVKGELKVLKQDKDTK 211
              |||||||||||||||||||||
Db        325 AGIEGTVKGELKVLKQDKDTK 345
```

RESULT 5
AAE11855
ID AAE11855 standard; protein; 168 AA.
XX
AC AAE11855;
XX
DT 18-DEC-2001 (first entry)
XX
DE Staphylococcus aureus CNA19 protein.
XX
KW Collagen-binding region; CNA19; Staphylococcus aureus infection;
KW Staphylococcus epidermidis infection; vaccine; CNA protein;
KW cross-reactive antibody.
XX
OS Staphylococcus aureus.
XX

FH	Key	Location/Qualifiers
FT	Region	24. .29
FT		/note= "Beta strand a; this region forms a part of the
FT		trench in the beta sheet"
FT	Region	37. .44
FT		/note= "Beta strand b; this region forms a part of the
FT		trench in the beta sheet"
FT	Region	55. .61
FT		/note= "Beta strand c"
FT	Region	65. .78
FT		/note= "Beta strand d; a portion of this region forms a
FT		part of the trench in the beta sheet"
FT	Region	82. .84
FT		/note= "Beta strand e; a portion of this region forms a
FT		part of the trench in the beta sheet"
FT	Region	89. .92
FT		/note= "Alpha helix 1"
FT	Region	93. .96
FT		/note= "Alpha helix 2"
FT	Region	101. .105
FT		/note= "Beta strand f"
FT	Region	110. .115
FT		/note= "Beta strand g"
FT	Region	123. .133
FT		/note= "Beta strand h; a portion of this region forms a
FT		part of the trench in the beta sheet"
FT	Region	140. .149
FT		/note= "Beta strand i"
FT	Region	157. .167
FT		/note= "Beta strand j"

XX
PN WO200170267-A1.
XX
PD 27-SEP-2001.
XX
PF 19-MAR-2001; 2001WO-US008554.
XX
PR 17-MAR-2000; 2000US-0189968P.
PR 25-APR-2000; 2000US-0199370P.
PR 15-AUG-2000; 2000US-0225402P.
XX
PA (INHI-) INHIBITEX INC.
PA (TEXA) UNIV TEXAS A & M SYSTEM.
PA (UYPA-) UNIV PAVIA.
XX
PI Hook M, Xu Y, Speziale P, Visai L, Casolini F, Patti J, Patel P;
PI Domanski P;
XX
DR WPI; 2001-607512/69.
XX
PT Novel isolated antibody which recognizes collagen-binding peptide such as
PT CNA19 peptide from Staphylococcus aureus, useful for preventing or
PT treating Staphylococcus aureus or epidermidis infection.
XX
PS Example 2; Fig 2A; 107pp; English.
XX
CC The invention relates to an antibody which recognises a collagen-binding
CC region including CNA19 of CNA protein from Staphylococcus aureus. This
CC antibody is cross-reactive to collagen binding region of both S. aureus
CC and S. epidermidis. It is useful for preventing or treating S. aureus or
CC S. epidermidis infection in human or animal, and for displacing S. aureus
CC or S. epidermidis bound to collagen. Antibody of the invention is useful
CC for interfering with, modulating, and inhibiting the binding interactions
CC between Staphylococcal bacteria and collagen, for detecting the presence
CC of Staphylococcal bacteria or Staphylococcal collagen or binding
CC proteins, to diagnose Staphylococcal infection, as research tools, for
CC development of vaccine for passive immunisation against Staphylococcal

CC infections, and in production facilities or laboratories to isolate
CC additional quantities of collagen-binding proteins. It is also useful for
CC treating medical instruments in order to reduce or eliminate the
CC possibility of them becoming infected or further spreading the infection.
CC The present sequence is S. aureus CNA19 protein
XX
SQ Sequence 168 AA;

Query Match 78.8%; Score 884; DB 4; Length 168;
Best Local Similarity 99.4%; Pred. No. 5.3e-66;
Matches 167; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 19 ITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSKDITIKDQ 78
Db 1 ITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSKDITIKDQ 60

Qy 79 IQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIPQGYG 138
Db 61 IQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIPQGYG 120

Qy 139 SYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHN 186
Db 121 SYNSFSINYKTKIINEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHN 168

RESULT 6
AAW31552
ID AAW31552 standard; protein; 159 AA.
XX
AC AAW31552;
XX
DT 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 21-MAY-1998 (first entry)
XX
DE Collagen binding protein M17 epitope.
XX
KW Collagen binding protein; cna gene; sepsis; infection;
KW microbial surface component regonising adhesive matrix molecule; MSCRAMM;
KW adhesin; vaccine; immunisation; diagnosis; therapy; epitope M17.
XX
OS Staphylococcus aureus.
XX
FH Key Location/Qualifiers
FT Peptide 1. .12
FT /note= "vector pQE30-derived peptide"
FT Protein 13. .159
FT /note= "epitope M17"
XX
PN WO9743314-A2.
XX
PD 20-NOV-1997.
XX
PF 14-MAY-1997; 97WO-US008210.
XX
PR 16-MAY-1996; 96US-0017678P.
XX
PA (TEXA) UNIV TEXAS A & M SYSTEM.
PA (UABR-) UAB RES FOUND.
XX
PI Hoeoek M, Patti JM, House-Pompeo K, Sthanam N, Symersky J;
XX
DR WPI; 1998-008801/01.
DR N-PSDB; AAT93436.
XX
PT Antibody that interacts with collagen binding domain of Staphylococcal
PT cna gene product - useful to prevent bacterial sepsis in animal infected
PT with Staphylococcus aureus.
XX
PS Claim 31; Page 114; 143pp; English.
XX
CC This protein comprises Staphylococcus aureus collagen binding protein
CC (CBP) epitope M17, i.e. amino acids 151-297 of full-length CBP, plus a
CC vector-derived N-terminal peptide. Claimed 441, 849 and 1500 bp nucleic
CC acid sequences (see AAT93436-38) respectively encode CBP epitopes M17,
CC M31 and M55 (see AAW31552-54) that confer protection against S. aureus
CC infection. These nucleic acid sequences can be used in the recombinant
CC production of the CBP epitopes. The CBP protein and antigenic epitopes
CC are contemplated for use in the treatment of pathological infections,
CC especially to prevent bacterial adhesion to collagen. The claimed nucleic
CC acids as well as claimed anti-CBP antibodies will also be of use in
CC screening, diagnostic and therapeutic applications including active and
CC passive immunisation and methods for the prevention of bacterial
CC colonisation in an animal such as a human. The CBP epitopes are also
CC contemplated for use in the preparation of vaccines and as carrier
CC proteins in vaccine formulations, as well as in the formulation of
CC compositions for the prevention of S. aureus infection. (Updated on 25-
CC MAR-2003 to correct PI field.) (Updated on 27-AUG-2003 to correct OS
CC field.)
XX
SQ Sequence 159 AA;

Query Match 74.0%; Score 830; DB 2; Length 159;
Best Local Similarity 96.4%; Pred. No. 1.6e-61;
Matches 159; Conservative 0; Mismatches 0; Indels 6; Gaps 1;

Qy 1 MRGSHHHHHHGSDDKVATITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFL 60
Db 1 MRGSHHHHHHGS------ITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFL 54

Qy 61 NINNEKSYVSKDITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKI 120
Db 55 NINNEKSYVSKDITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKI 114

Qy 121 TVDNTKNTIDVTIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQA 165
Db 115 TVDNTKNTIDVTIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQA 159

RESULT 7
ADC97372
ID ADC97372 standard; protein; 731 AA.
XX
AC ADC97372;
XX
DT 01-JAN-2004 (first entry)
XX
DE E. faecium protein sequence SEQ ID 6999.
XX
KW Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;
KW abdominal-pelvic infection.
XX
OS Enterococcus faecium.
XX
PN US6583275-B1.
XX
PD 24-JUN-2003.
XX
PF 30-JUN-1998; 98US-00107532.
XX
PR 02-JUL-1997; 97US-0051571P.
PR 14-MAY-1998; 98US-0085598P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
DR WPI; 2003-799836/75.
DR N-PSDB; ADC93718.
XX
PT New isolated nucleic acid derived from Enterococcus faecium encoding an
PT Enterococcus faecium polypeptide useful for detection, prevention and
PT treatment of a pathological condition resulting from a bacterial
PT infection.
XX
PS Example 1; SEQ ID NO 6999; 243pp; English.

```
XX
CC   The invention relates to an isolated nucleic acid derived from
CC   Enterococcus faecium encoding an Enterococcus faecium polypeptide having
CC   one of 10 fully defined sequences given in the (or comprising 40
CC   sequential nucleotides chosen from any of the nucleic acids, its
CC   complement or sequences hybridising to it). Also included are a
CC   recombinant vector comprising the nucleic acid operably linked to
CC   transcription regulatory element, a cell comprising the vector and a
CC   single-stranded probe comprising the nucleic acid. The nucleic acids are
CC   chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
CC   The nucleic acids is useful for diagnosing pathological conditions
CC   resulting from E. faecium bacterial infection (e.g. urinary tract
CC   infection, bacteraemia, endocarditis, wounds and abdominal-pelvic
CC   infection) and for screening drugs such as agonists and antagonists. The
CC   nucleic acid is useful for recombinant production of Candida albicans -
CC   derived peptides or antisense polypeptides. Pharmaceutical compositions
CC   and vaccines containing the nucleic acid are useful for preventing or
CC   treating Enterococcus faecium infections. The present sequence represents
CC   one if the disclosed E. faecium proteins.
XX
SQ   Sequence 731 AA;

  Query Match             43.8%;  Score 491;  DB 7;  Length 731;
  Best Local Similarity   48.5%;  Pred. No. 2.6e-32;
  Matches   95;  Conservative   35;  Mismatches   64;  Indels    2;  Gaps    2;

Qy         18 TITSGNKSTNVTVHKSEAG-TSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSKDITIK 76
              |:|||:|:  | | |  :| :|||||||||||||||| |:||||||||  :|| : : |
Db        162 TVTSGDKTATVNVTKPASGSSSSVFYYKTGDMLPEDTKHIRWFLNINNNGTYVEQPVKIS 221

Qy         77 DQIQGGQQLDLSTLNIN-VTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIPQ 135
              |:|| ||:|| ||  || :       | |:  |  | : || :       | | |::|||:
Db        222 DEIQSGQRLDPSTFEINQIHLGEQKVYRGEEGIQQFLQDFPSATFNFSVTDNYIEITIPK 281

Qy        136 GYGSYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINANAGIEG 195
               : :     ::||| | | :|  | |:|:||::|  |  |:|:|||||| ||:|: |: |
Db        282 NFVNLRKIMVSYKTIIENPEQINFENHSEAWFKEFNKPAVDGESFNHTVKNISASGGVNG 341

Qy        196 TVKGELKVLKQDKDTK 211
              ||:||||: |   ||:
Db        342 TVRGELKIFKYINDTE 357


RESULT 8
AAB37667
ID   AAB37667 standard; protein; 458 AA.
XX
AC   AAB37667;
XX
DT   28-FEB-2001  (first entry)
XX
DE   Collagen binding Ace protein.
XX
KW   Collagen binding Ace protein; antibacterial; immunostimulant; vaccine;
KW   extracellular matrix-binding protein; lamin binding; bacterial infection.
XX
OS   Enterococcus faecalis.
XX
PN   WO200068242-A1.
XX
PD   16-NOV-2000.
XX
PF   10-MAY-2000; 2000WO-US012590.
XX
PR   10-MAY-1999;   99US-0133334P.
XX
PA   (TEXA ) UNIV TEXAS A & M SYSTEM.
PA   (TEXA ) UNIV TEXAS MEDICAL SCHOOL.
XX
PI   Rich RL,  Kriekemeyer B,  Owens RT,  Hook M,  Murray BE;
PI   Nallapareddy SR,  Qin X,  Weinstock GM,  Singh KV,  Duh R;
XX
DR   WPI; 2000-687639/67.
DR   N-PSDB; AAC83828.
XX
PT   New collagen-binding protein from Enterococcus, useful e.g. in protective
PT   vaccines, for diagnosis and treatment of Enterococcal infections and for
PT   screening for compounds that inhibit collagen binding by enterococci.
XX
PS   Claim 26; Page 119-122; 148pp; English.
XX
CC   The present sequence is collagen binding Ace protein from Enterococcus
CC   faecalis, where Ace is adhesion of collagen from enterococcal bacteria.
CC   This protein is an extracellular matrix-binding protein, which can bind
CC   with collagens such as collagen type 1 and type IV and with lamins. Ace
CC   protein can be used to inhibit attachment of enterococci to collagen and
CC   to treat or prevent enterococcal infections, specifically as a vaccine
XX
SQ   Sequence 458 AA;

  Query Match             15.9%;  Score 178;  DB 3;  Length 458;
  Best Local Similarity   28.3%;  Pred. No. 2.1e-06;
  Matches   54;  Conservative   33;  Mismatches   88;  Indels   16;  Gaps    6;

Qy         18 TITSGNKSTNVTVHKSEAG---TSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSKDIT 74
              | |:  : |   |  :| |    .  |:|| ||:  | :  ||||||:|  || |::||:
Db        121 TATATQRLTIEGVTNTETGQIERDYPFFYKVGDLAGE-SNQVRWFLNVNLNKSDVTEDIS 179

Qy         75 IKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIP 134
              | |:   ||||:  :   ::        |    :: :||:   |    |  | |  ::
Db        180 IADRQGSGQQLNKESFTFDIVNDKETKY---ISLAEFEQQGYGKIDFV--TDNDFNLRFY 234

Qy        135 QGYGSYNSFSINYKTKIT--NEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINANAG 192
              :    : || : | : ||   : |  | |:    || : ::  | |: :        |
Db        235 RDKARFTSFIVRYTSTITEAGQHQATFENSYDINYQLNNQDATNEKNTSQV-----KNVF 289

Qy        193 IEGTVKGELKV 203
              :||   |   |
Db        290 VEGEASGNQNV 300


RESULT 9
ADD90648
ID   ADD90648 standard; protein; 458 AA.
XX
AC   ADD90648;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Enterococcus faecalis collagen binding ACE protein SEQ ID NO:1.
XX
KW   cross-reactive monoclonal antibody; ACE; Enterococcus faecalis;
KW   antibacterial; veterinary; antibody therapy; vaccine; infection;
KW   collagen binding ACE.
XX
OS   Enterococcus faecalis.
XX
PN   WO2003072607-A1.
XX
PD   04-SEP-2003.
XX
PF   21-FEB-2003; 2003WO-US005040.
XX
PR   21-FEB-2002; 2002US-0357832P.
PR   05-MAR-2002; 2002US-0361347P.
XX
PA   (UYPA-) UNIV PAVIA.
XX
PI   Bates SE,  Speziale P,  Visai L,  Giampiero P;
XX
DR   WPI; 2003-679944/64.
DR   N-PSDB; ADD90649.
XX
PT   New cross-reactive monoclonal antibody that binds to a peptide such as
PT   the A domain of the ACE protein from Enterococcus faecali, or CAN19 from
PT   Staphylococcus aureus, useful for treating or preventing infections
```

PT caused by the organisms.
XX
PS Claim 8; SEQ ID NO 1; 49pp; English.
XX
CC The present invention describes a cross-reactive monoclonal antibody (I)
CC that binds to a peptide selected from the A domain of the ACE protein
CC from Enterococcus faecalis, ACE 19 from E. faecalis and CAN19 from
CC Staphylococcus aureus. Also described: (1) an isolated antisera
CC containing (I); (2) a diagnostic kit comprising (I) and means for
CC detecting binding by (I); (3) a pharmaceutical composition for treating
CC or preventing a bacterial infection comprising (I), and a vehicle,
CC carrier or excipient; (4) a method for treating or preventing an
CC infection of enterococcal, streptococcal, or staphylococcal infection by
CC administering (I) to a human or animal patient; and (5) a method for
CC inducing an immmonological response by administering to human or animal
CC an immunogenic amount of an isolated protein from E. faecalis selected
CC from ACE40 and ACE19 protein. (I) has antibacterial and veterinary
CC activities, and can be used in antibody therapy, and in vaccines. The
CC antibody and pharmaceutical composition are useful for treating or
CC preventing infection from staphylococcal, streptococcal and enterococcal
CC bacteria in a human or animal. The present sequence represents the
CC Enterococcus faecalis collagen binding ACE protein, which is used in the
CC exemplification of the present invention.
XX
SQ Sequence 458 AA;

Query Match 15.9%; Score 178; DB 7; Length 458;
Best Local Similarity 28.3%; Pred. No. 2.1e-06;
Matches 54; Conservative 33; Mismatches 88; Indels 16; Gaps 6;

```
Qy          18 TITSGNKSTNVTVHKSEAG---TSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSKDIT 74
               | |:  : |   |  :| |       |:|| ||:  | :  ||||||:|  || |::||:
Db         121 TATATQRLTIEGVTNTETGQIERDYPFFYKVGDLAGE-SNQVRWFLNVNLNKSDVTEDIS 179

Qy          75 IKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIP 134
               | |:   ||||:  :   ::       |    :: :||:   |    |  | |  ::
Db         180 IADRQGSGQQLNKESFTFDIVNDKETKY---ISLAEFEQQGYGKIDFV--TDNDFNLRFY 234

Qy         135 QGYGSYNSFSINYKTKIT--NEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINANAG 192
               :    : || : | : ||   : |  | |:     || : ::  | |: :         |
Db         235 RDKARFTSFIVRYTSTITEAGQHQATFENSYDINYQLNNQDATNEKNTSQV-----KNVF 289

Qy         193 IEGTVKGELKV 203
               :||   |   |
Db         290 VEGEASGNQNV 300
```

RESULT 10
AAB37668
ID AAB37668 standard; protein; 146 AA.
XX
AC AAB37668;
XX
DT 28-FEB-2001 (first entry)
XX
DE Collagen binding Ace protein collagen binding domain.
XX
KW Collagen binding Ace protein; antibacterial; immunostimulant; vaccine;
KW extracellular matrix-binding protein; lamin binding; bacterial infection.
XX
OS Enterococcus faecalis.
XX
PN WO200068242-A1.
XX
PD 16-NOV-2000.
XX
PF 10-MAY-2000; 2000WO-US012590.
XX
PR 10-MAY-1999; 99US-0133334P.
XX
PA (TEXA) UNIV TEXAS A & M SYSTEM.
PA (TEXA) UNIV TEXAS MEDICAL SCHOOL.
XX
PI Rich RL, Kriekemeyer B, Owens RT, Hook M, Murray BE;
PI Nallapareddy SR, Qin X, Weinstock GM, Singh KV, Duh R;
XX
DR WPI; 2000-687639/67.
XX
PT New collagen-binding protein from Enterococcus, useful e.g. in protective
PT vaccines, for diagnosis and treatment of Enterococcal infections and for
PT screening for compounds that inhibit collagen binding by enterococci.
XX
PS Disclosure; Fig 1; 148pp; English.
XX
CC The present invention relates to collagen binding Ace protein from
CC Enterococcus faecalis, where Ace is adhesion of collagen from
CC enterococcal bacteria (see AAC83828 and AAB37667). Ace protein is an
CC extracellular matrix-binding protein, which can bind with collagens such
CC as collagen type 1 and type IV and with lamins. Ace protein can be used
CC to inhibit attachment of enterococci to collagen and to treat or prevent
CC enterococcal infections, specifically as a vaccine. The present sequence
CC is the collagen binding domain of Ace protein
XX
SQ Sequence 146 AA;

Query Match 14.8%; Score 165.5; DB 3; Length 146;
Best Local Similarity 30.2%; Pred. No. 5e-06;
Matches 45; Conservative 30; Mismatches 65; Indels 9; Gaps 5;

```
Qy          41 FYYKTGDMLPEDTTHVRWFLNINNEKSYVSKDITIKDQIQGGQQLDLSTLNINVTGTHSN 100
               |:|| ||:  | :  ||||||:|  || |::||:| |:   ||||:  :   ::
Db           5 FFYKVGDLAGE-SNQVRWFLNVNLNKSDVTEDISIADRQGSGQQLNKESFTFDIVNDKET 63

Qy         101 YYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIPQGYGSYNSFSINYKTKIT--NEQQKE 158
               |    :: :||:   |    |  | |  ::   :    : || : | : ||   : |
Db          64 KY---ISLAEFEQQGYGKIDFV--TDNDFNLRFYRDKARFTSFIVRYTSTITEAGQHQAT 118

Qy         159 FVNNSQAWYQEHGKEEVNGKSFNHTVHNI 187
               | |:   || : ::  | |: :  | |:
Db         119 FENSYDINYQLNNQDATNEKNTSQ-VKNV 146
```

RESULT 11
AAE11856
ID AAE11856 standard; protein; 168 AA.
XX
AC AAE11856;
XX
DT 18-DEC-2001 (first entry)
XX
DE Enterococcus faecalis ACE19 protein.
XX
KW Collagen-binding region; CNA19; Staphylococcus aureus infection;
KW Staphylococcus epidermidis infection; vaccine; CNA protein;
KW cross-reactive antibody; collagen-adhesin; ACE.
XX
OS Enterococcus faecalis.
XX
PN WO200170267-A1.
XX
PD 27-SEP-2001.
XX
PF 19-MAR-2001; 2001WO-US008554.
XX
PR 17-MAR-2000; 2000US-0189968P.
PR 25-APR-2000; 2000US-0199370P.
PR 15-AUG-2000; 2000US-0225402P.
XX
PA (INHI-) INHIBITEX INC.
PA (TEXA) UNIV TEXAS A & M SYSTEM.
PA (UYPA-) UNIV PAVIA.
XX
PI Hook M, Xu Y, Speziale P, Visai L, Casolini F, Patti J, Patel P;
PI Domanski P;
XX
DR WPI; 2001-607512/69.

```
XX
PT   Novel isolated antibody which recognizes collagen-binding peptide such as
PT   CNA19 peptide from Staphylococcus aureus, useful for preventing or
PT   treating Staphylococcus aureus or epidermidis infection.
XX
PS   Example 2; Fig 2A; 107pp; English.
XX
CC   The invention relates to an antibody which recognises a collagen-binding
CC   region including CNA19 of CNA protein from Staphylococcus aureus. This
CC   antibody is cross-reactive to collagen binding region of both S. aureus
CC   and S. epidermidis. It is useful for preventing or treating S. aureus or
CC   S. epidermidis infection in human or animal, and for displacing S. aureus
CC   or S. epidermidis bound to collagen. Antibody of the invention is useful
CC   for interfering with, modulating, and inhibiting the binding interactions
CC   between Staphylococcal bacteria and collagen, for detecting the presence
CC   of Staphylococcal bacteria or Staphylococcal collagen or binding
CC   proteins, to diagnose Staphylococcal infection, as research tools, for
CC   development of vaccine for passive immunisation against Staphylococcal
CC   infections, and in production facilities or laboratories to isolate
CC   additional quantities of collagen-binding proteins. It is also useful for
CC   treating medical instruments in order to reduce or eliminate the
CC   possibility of them becoming infected or further spreading the infection.
CC   The present sequence is Enterococcus faecalis collagen adhesion (ACE)19
CC   protein. This sequence is homologous to S. aureus CNA19 protein
XX
SQ   Sequence 168 AA;

  Query Match             13.7%;  Score 154;  DB 4;  Length 168;
  Best Local Similarity   25.8%;  Pred. No. 5.5e-05;
  Matches   50;  Conservative   31;  Mismatches   63;  Indels   50;  Gaps    8;

Qy         18 TITSGNKSTNVTVHKSEAG---TSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSKDIT 74
              | |:  : |   |  :| |       |:|| ||:  | :  |||||:|  || |::||:
Db          1 TATATQRLTIEGVTNTETGQIERDYPFFYKVGDLAGE-SNQVRWFLNVNLNKSDVTEDIS 59

Qy         75 IKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIP 134
              | |:   |||::  :  ::      |    :: :||:
Db         60 IADRQGSGQQINKESFTFDIVNDKETKY---ISLAEFEQ---------------------- 95

Qy        135 QGYGSY-----NSFS-------------INYKTKIT--NEQQKEFVNNSQAWYQEHGKE 173
              ||||       | |:            : | : ||   : |  | |:    || : ::
Db         96 QGYGKIDFVTDNDFNLRFYRDKARETSEIVRYTSTITEAGQHQATFENSYDINYQLNNQD 155

Qy        174 EVNGKSFNHTVHNI 187
                | |: :  | |:
Db        156 ATNEKNTSQ-VKNV 168


RESULT 12
ABM79017
ID   ABM79017 standard; protein; 316 AA.
XX
AC   ABM79017;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Staphylococcus epidermidis SdrG TR2 protein.
XX
KW   SdrG; surface protein; infection; antibacterial; vaccine.
XX
OS   Staphylococcus epidermidis.
XX
PN   WO2003076470-A1.
XX
PD   18-SEP-2003.
XX
PF   05-MAR-2003; 2003WO-US006415.
XX
PR   05-MAR-2002; 2002US-0361324P.
XX
PA   (INHI-) INHIBITEX INC.
PA   (TEXA ) UNIV TEXAS A & M SYSTEM.
XX
PI   Patti JM,  Hutchins JT,  Hall A,  Domanski P,  Patel P,  Hook M;
PI   Robbins J,  Vernachio J,  Bowden MG;
XX
DR   WPI; 2003-722324/68.
DR   N-PSDB; ACF80626.
XX
PT   New antibody recognizing a Staphylococcus epidermidis protein comprising
PT   SdrG N1N2N3, SdrG N2N3 or SdrGTR2 useful for preparing a composition for
PT   treating or preventing a coagulase-negative Staphylococcal infection.
XX
PS   Claim 27; Page 26; 78pp; English.
XX
CC   The present sequence comprises the protein sequence of the TR2 protein
CC   (amino acids 273-577) of the SdrG surface protein of coagulase-negative
CC   Staphylococcus epidermidis. A claimed antibody recognises a protein
CC   selected from SdrG N1N2N3, N2N3 and TR2. The antibody may be a monoclonal
CC   antibody, including a chimeric, murine, humanized, human or single chain
CC   monoclonal antibody, which prevents a coagulase-negative staphylococcal
CC   infection in a human or animal by inhibiting binding of staphylococcus
CC   bacteria to fibrinogen. Such antibodies can be used to treat or prevent
CC   staphylococcal infections including nosocomial coagulase-negative
CC   staphylococcal infections in low birth weight infants. A claimed vaccine
CC   comprises the isolated SdrG N1N2N3, N2N3 or TR2 protein
XX
SQ   Sequence 316 AA;

  Query Match             12.6%;  Score 141.5;  DB 7;  Length 316;
  Best Local Similarity   26.5%;  Pred. No. 0.0014;
  Matches   67;  Conservative   32;  Mismatches   79;  Indels   75;  Gaps   16;

Qy          1 MRGSHHHHHHGSDD--------KVA--TITSGNKSTN--VTVHKSEAGTSSVFY---YKT 45
              ||||||||||||:          ||   :|| |   ::  :  | :|    :: |   ::
Db          1 MRGSHHHHHHGSEQGSNVNHLIKVTDQSITEGYDDSDGIIKAHDAE----NLIYDVTFEV 56

Qy         46 GDMLPEDTTHVRWFLNINNEKSYVSKDIT-------IKDQIQGGQQLDLST---LNINVT 95
               | :    |     : :| :|: |  |:|       |||    |: :   |    |  :|
Db         57 DDKVKSGDT-----MTVNIDKNTVPSDLTDSFAIPKIKD--NSGEIIATGTYDNTNKQIT 109

Qy         96 GTHSNY---YSGQSAITDFEKAFPGSKITVDNTKNTIDVTIPQGYGSYNSFSINYKTKIT 152
              | ::|   |   |          ||:  :|||  :||       | |      || ||
Db        110 YTFTDYVDKYENIKAHLKLTSYIDKSKVPNNNTK--LDVEYKTALSSVN------KT-IT 160

Qy        153 NEQQKEFVN---NSQAWYQEHGKEEVNGKSFNHTVH-----------------NINANAG 192
              | ||   |   |:|: :        |  : ||||                  ||: |
Db        161 VEYQKPNENRTANLQSMF-------TNIDTKNHTVEQTIYINPLRYSAKETNVNISGNGD 213

Qy        193 IEGTVKGELKVLK 205
                |:  :  ::|
Db        214 EGSTIIDDSTIIK 226


RESULT 13
ABM79016
ID   ABM79016 standard; protein; 343 AA.
XX
AC   ABM79016;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Staphylococcus epidermidis SdrG N2N3 domain.
XX
KW   SdrG; surface protein; infection; antibacterial; vaccine.
XX
OS   Staphylococcus epidermidis.
XX
PN   WO2003076470-A1.
XX
PD   18-SEP-2003.
XX
PF   05-MAR-2003; 2003WO-US006415.
XX
PR   05-MAR-2002; 2002US-0361324P.
XX
```

```
PA   (INHI-) INHIBITEX INC.
PA   (TEXA ) UNIV TEXAS A & M SYSTEM.
XX
PI   Patti JM,  Hutchins JT,  Hall A,  Domanski P,  Patel P,  Hook M;
PI   Robbins J,  Vernachio J,  Bowden MG;
XX
DR   WPI; 2003-722324/68.
DR   N-PSDB; ACF80625.
XX
PT   New antibody recognizing a Staphylococcus epidermidis protein comprising
PT   SdrG N1N2N3, SdrG N2N3 or SdrGTR2 useful for preparing a composition for
PT   treating or preventing a coagulase-negative Staphylococcal infection.
XX
PS   Claim 27; Page 25; 78pp; English.
XX
CC   The present sequence comprises the protein sequence of the N2N3 region
CC   (amino acids 273-597) of the SdrG surface protein of coagulase-negative
CC   Staphylococcus epidermidis. A claimed antibody recognises a protein
CC   selected from SdrG N1N2N3, N2N3 and TR2. The antibody may be a monoclonal
CC   antibody, including a chimeric, murine, humanized, human or single chain
CC   monoclonal antibody, which prevents a coagulase-negative staphylococcal
CC   infection in a human or animal by inhibiting binding of staphylococcus
CC   bacteria to fibrinogen. Such antibodies can be used to treat or prevent
CC   staphylococcal infections including nosocomial coagulase-negative
CC   staphylococcal infections in low birth weight infants. A claimed vaccine
CC   comprises the isolated SdrG N1N2N3, N2N3 or TR2 protein
XX
SQ   Sequence 343 AA;

  Query Match             11.7%;  Score 131.5;  DB 7;  Length 343;
  Best Local Similarity   25.3%;  Pred. No. 0.011;
  Matches   66;  Conservative   32;  Mismatches   78;  Indels   85;  Gaps   16;

Qy          1 MRGSHHHHHHGS------------------DDKVATITSGNKSTN--VTVHKSEAGTSSV 40
              ||||||||||||                  |:  :|| |   ::  :  | :|    ::
Db          1 MRGSHHHHHHGSLVPRGSEQGSNVNHLIKVTDQ--SITEGYDDSDGIIKAHDAE----NL 54

Qy         41 FY---YKTGDMLPEDTTHVRWFLNINNEKSYVSKDIT-------IKDQIQGGQQLDLST- 89
               |   ::  | :    |     : :| :|: |  |:|       |||    |: :   |
Db         55 IYDVTFEVDDKVKSGDT-----MTVNIDKNTVPSDLTDSFAIPKIKD--NSGEIIATGTY 107

Qy         90 --LNINVTGTHSNY---YSGQSAITDFEKAFPGSKITVDNTKNTIDVTIPQGYGSYNSFS 144
                |  :| | ::|   |    |           ||:  :||| :||       | |
Db        108 DNTNKQITYTFTDYVDKYENIKAHLKLTSYIDKSKVPNNNTK--LDVEYKTALSSVN--- 162

Qy        145 INYKTKITNEQQKEFVN---NSQAWYQEHGKEEVNGKSFNHTVH---------------- 185
              || || | ||   |   | |: :       |  : ||||
Db        163 ---KT-ITVEYQKPNENRTANLQSMF-------TNIDTKNHTVEQTIYINPLRYSAKETN 211

Qy        186 -NINANAGIEGTVKGELKVLK 205
               ||: |     |:  :  ::|
Db        212 VNISGNGDEGSTIIDDSTIIK 232


RESULT 14
AAW31555
ID   AAW31555 standard; protein; 345 AA.
XX
AC   AAW31555;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   21-MAY-1998  (first entry)
XX
DE   Firbonectin-binding MSCRAMM derivative pCF33.
XX
KW   Fibronectin; pCF33; collagen binding protein; sepsis; infection;
KW   microbial surface component regonising adhesive matrix molecule; MSCRAMM;
KW   adhesin; vaccine; immunisation; diagnosis; therapy; mastitis.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .12
FT                   /note= "vector pQE30-derived peptide"
XX
PN   WO9743314-A2.
XX
PD   20-NOV-1997.
XX
PF   14-MAY-1997;   97WO-US008210.
XX
PR   16-MAY-1996;   96US-0017678P.
XX
PA   (TEXA ) UNIV TEXAS A & M SYSTEM.
PA   (UABR-) UAB RES FOUND.
XX
PI   Hoeoek M,  Patti JM,  House-Pompeo K,  Sthanam N,  Symersky J;
XX
DR   WPI; 1998-008801/01.
XX
PT   Antibody that interacts with collagen binding domain of Staphylococcal
PT   cna gene product - useful to prevent bacterial sepsis in animal infected
PT   with Staphylococcus aureus.
XX
PS   Disclosure; Page 91; 143pp; English.
XX
CC   This protein comprises Staphylococcus aureus fibronectin-binding
CC   microbial surface component regonising adhesive matrix molecule (MSCRAMM)
CC   derivative pCF33, plus a vector-derived N-terminal peptide. The invention
CC   relates to claimed nucleic acid sequences (see AAT93436-38) encoding S.
CC   aureus collagen binding protein (CBP) epitopes M17, M31 and M55 (see
CC   AAW31552-54) that confer protection against S. aureus infection. CBP
CC   protein and antigenic epitopes are contemplated for use in the treatment
CC   of pathological infections, especially to prevent bacterial adhesion to
CC   collagen. The epitopes are also contemplated for use in the preparation
CC   of vaccines and as carrier proteins in vaccine formulations, as well as
CC   in the formulation of compositions for the prevention of S. aureus
CC   infection. pCF33 and pQD (see AAW31556) were used to raise anti-MSCRAMM
CC   polyclonal antibodies used in passive immunisation against bovine
CC   mastitis. (Updated on 25-MAR-2003 to correct PI field.) (Updated on 27-
CC   AUG-2003 to correct OS field.)
XX
SQ   Sequence 345 AA;

  Query Match             11.2%;  Score 125.5;  DB 2;  Length 345;
  Best Local Similarity   30.1%;  Pred. No. 0.035;
  Matches   56;  Conservative   20;  Mismatches   73;  Indels   37;  Gaps   10;

Qy          1 MRGSHHHHHHGS----DDKVATITSGNKSTNV--------TVHKSEAGTSSVFYYKTGDM 48
              ||||||||||||    |    |      |: |||        ||:  :||   : |   |
Db          1 MRGSHHHHHHGSMVAADAPAAGTDITNQLTNVTVGIDSGTTVYPHQAGYVKLNY---GFS 57

Qy         49 LPEDTTHVRWF-------LNINNEKSYVS-KDITIKDQIQGGQQLDLSTLNINVTGTHSN 100
              :|          |       ||:|   |       |   ||:     :|      ||  | ::
Db         58 VPNSAVKGDTFKITVPKELNLNGVTSTAKVPPIMAGDQVLANGVIDSDG---NVIYTFTD 114

Qy        101 YYSGQSAITDFEKA---FPGSKITVDNTKNTIDVTIPQGYGSYNSFSINYKTKITNEQQK 157
              | : :    |  ||    | : |  :| | | :||:  | ||  :   |    :  |:
Db        115 YVNTK----DDVKATLTMP-AYIDPENVKKTGNVTLATGIGSTTA---NKTVLVDYEKYG 166

Qy        158 EFVNNS 163
              :| | |
Db        167 KFYNLS 172


RESULT 15
ABB53995
ID   ABB53995 standard; protein; 336 AA.
XX
AC   ABB53995;
XX
DT   29-AUG-2003  (revised)
DT   16-MAY-2002  (first entry)
XX
DE   Lactococcus lactis protein icaC.
```

XX
KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX
OS Lactococcus lactis; IL1403.
XX
PN FR2807446-A1.
XX
PD 12-OCT-2001.
XX
PF 11-APR-2000; 2000FR-00004630.
XX
PR 11-APR-2000; 2000FR-00004630.
XX
PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
XX
PI Bolotine A, Sorokine A, Renault P, Ehrlich SD;
XX
DR WPI; 2002-043418/06.
XX
PT New nucleotide sequence useful in the identification or Lactococcus
PT lactis and related species.
XX
PS Claim 6; SEQ ID NO 697; 2504pp; French.
XX
CC The present invention is related to a Lactococcus lactis nucleotide
CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The nucleic
CC acid sequence is useful in the detection and/or amplification of nucleic
CC acid sequence, particularly to identify Lactococcus lactis or related
CC species. The proteins of the invention are useful for the biosynthesis or
CC biodegradation of a composition of interest. The invention helps research
CC in lactic bacteria, particularly useful in the production of yogurt and
CC cheese. Note: The sequence data for this patent is based on equivalent
CC patent WO200177334 (published 18-OCT-2001) which is available in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences. (Updated on 29-AUG-2003 to
CC standardise OS field)
XX
SQ Sequence 336 AA;

Query Match 10.6%; Score 119; DB 5; Length 336;
Best Local Similarity 25.2%; Pred. No. 0.12;
Matches 52; Conservative 28; Mismatches 68; Indels 58; Gaps 11;

```
Qy          24 KSTNVTVHKSEAGTSSV------------FYYKTGDMLPEDTTHVRWFLNINNEKSYVSK 71
               : | | |   : ||: |               || | :   |:| : | : ||  |  : |
Db          71 EDTRVVVDWKDGGTTEVNIGPATGPDKDEVLYKWGWVDENDSTLIHWQVRINYAKENIQK 130

Qy          72 DITIKDQIQGGQQL---DLSTLNINVTGTHSNY----YSGQSAITDFEKAFPGSKITVDN 124
                 |   | | | | |    :|  |:  :    ||  .  |  |::|   |    |  : : :
Db         131 AI-YTDIIGGNQNLVSGSISVANVTYSSDGENYNVDSYYPQASI--LENGVNGFTVNLGD 187

Qy         125 TKNTIDVTIPQGYGSYNSFSINYKTKITN---EQQKEFVNNSQAWYQEHGKEEVNGKSFN 181
                 |||              |:| |: |:    ||          |:  |  |: |::
Db         188 ISNTI-------------IIDYSTRATDGGLSQQ----------YENRG--ELTGENIE 221

Qy         182 ------HTVHN-INANAGIEGTVKGE 200
                    || :|  | || :: :: ||
Db         222 KQVVEVHTPNNGGNGNASMKLSISGE 247
```

Search completed: July 27, 2004, 10:08:19
Job time : 57 secs

OM protein - protein search, using sw model

Run on: July 27, 2004, 10:07:21 ; Search time 19 Seconds
(without alignments)
573.320 Million cell updates/sec

Title: US-09-813-820-4
Perfect score: 1122
Sequence: 1 MRGSHHHHHHGSDDKVATIT.....GIEGTVKGELKVLKQDKDTK 211

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1122	100.0	211	3	US-08-856-253-4	Sequence 4, Appli
2	1054.5	94.0	512	3	US-08-856-253-6	Sequence 6, Appli
3	1048	93.4	1183	2	US-08-447-031A-2	Sequence 2, Appli
4	830	74.0	159	3	US-08-856-253-2	Sequence 2, Appli
5	491	43.8	731	4	US-09-107-532A-6999	Sequence 6999, Ap
6	178	15.9	540	4	US-09-134-000C-3745	Sequence 3745, Ap
7	125.5	11.2	345	3	US-08-856-253-7	Sequence 7, Appli
8	114.5	10.2	688	4	US-09-071-035-464	Sequence 464, App
9	114.5	10.2	2032	4	US-09-071-035-458	Sequence 458, App
10	114.5	10.2	2032	4	US-09-071-035-462	Sequence 462, App
11	114.5	10.2	2032	4	US-09-071-035-466	Sequence 466, App
12	114.5	10.2	2054	4	US-09-134-000C-6612	Sequence 6612, Ap
13	113.5	10.1	184	3	US-09-023-082A-30	Sequence 30, Appl
14	113.5	10.1	184	4	US-09-248-998-30	Sequence 30, Appl
15	110.5	9.8	1338	2	US-08-728-470-9	Sequence 9, Appli
16	110.5	9.8	1338	3	US-08-719-641-9	Sequence 9, Appli
17	110.5	9.8	1599	2	US-08-617-697-9	Sequence 9, Appli
18	110	9.8	2314	4	US-09-268-347-49	Sequence 49, Appl
19	104.5	9.3	1529	2	US-08-728-470-10	Sequence 10, Appl
20	104.5	9.3	1529	3	US-08-719-641-10	Sequence 10, Appl
21	104.5	9.3	1600	2	US-08-617-697-10	Sequence 10, Appl
22	103.5	9.2	139	3	US-08-856-253-8	Sequence 8, Appli
23	103.5	9.2	523	4	US-09-543-681A-5151	Sequence 5151, Ap
24	100	8.9	1095	4	US-09-206-942-69	Sequence 69, Appl
25	100	8.9	1536	1	US-08-038-682-2	Sequence 2, Appli
26	100	8.9	1536	1	US-08-302-832-2	Sequence 2, Appli
27	100	8.9	1536	2	US-08-530-198-2	Sequence 2, Appli
28	100	8.9	1536	2	US-08-469-880-2	Sequence 2, Appli
29	100	8.9	1536	2	US-08-728-470-2	Sequence 2, Appli
30	100	8.9	1536	2	US-08-617-697-2	Sequence 2, Appli
31	100	8.9	1536	3	US-08-719-641-2	Sequence 2, Appli
32	100	8.9	1536	4	US-09-206-942-67	Sequence 67, Appl
33	99.5	8.9	589	3	US-09-377-155-19	Sequence 19, Appl
34	99.5	8.9	589	4	US-09-669-974-19	Sequence 19, Appl
35	99.5	8.9	589	4	US-09-797-862-19	Sequence 19, Appl
36	98.5	8.8	351	2	US-08-701-191A-3	Sequence 3, Appli
37	98.5	8.8	351	4	US-09-664-526-3	Sequence 3, Appli
38	96.5	8.6	1742	4	US-09-386-962C-4	Sequence 4, Appli
39	95.5	8.5	1436	4	US-10-080-505-13	Sequence 13, Appl
40	95	8.5	666	2	US-08-737-716-14	Sequence 14, Appl
41	94.5	8.4	514	4	US-09-107-532A-4411	Sequence 4411, Ap
42	94	8.4	287	4	US-09-543-681A-5491	Sequence 5491, Ap
43	94	8.4	821	4	US-09-556-877-195	Sequence 195, App
44	94	8.4	821	4	US-09-620-412C-195	Sequence 195, App
45	94	8.4	821	4	US-09-598-419-195	Sequence 195, App

ALIGNMENTS

RESULT 1
US-08-856-253-4
; Sequence 4, Application US/08856253
; Patent No. 6288214
; GENERAL INFORMATION:
; APPLICANT: Hook, Magnus
; APPLICANT: Patti, Joseph M.
; APPLICANT: House-Pompeo, Karen
; APPLICANT: Sthanam, Narayana
; APPLICANT: Symersky, Jindrich
; TITLE OF INVENTION: COLLAGEN BINDING PROTEIN COMPOSITIONS
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/856,253
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/017,678
; FILING DATE: 16-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kitchell, Barbara S.
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: TAMK:193
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 211 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-856-253-4

Query Match 100.0%; Score 1122; DB 3; Length 211;
Best Local Similarity 100.0%; Pred. No. 4.7e-94;
Matches 211; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MRGSHHHHHHGSDDKVATITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRGSHHHHHHGSDDKVATITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFL 60

Qy         61 NINNEKSYVSKDITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NINNEKSYVSKDITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKI 120

Qy        121 TVDNTKNTIDVTIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TVDNTKNTIDVTIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSF 180

Qy        181 NHTVHNINANAGIEGTVKGELKVLKQDKDTK 211
              |||||||||||||||||||||||||||||||
Db        181 NHTVHNINANAGIEGTVKGELKVLKQDKDTK 211


RESULT 2
US-08-856-253-6
; Sequence 6, Application US/08856253
; Patent No. 6288214
;  GENERAL INFORMATION:
;    APPLICANT:  Hook, Magnus
;    APPLICANT:  Patti, Joseph M.
;    APPLICANT:  House-Pompeo, Karen
;    APPLICANT:  Sthanam, Narayana
;    APPLICANT:  Symersky, Jindrich
;    TITLE OF INVENTION:  COLLAGEN BINDING PROTEIN COMPOSITIONS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  U.S.
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/856,253
;      FILING DATE:  Concurrently Herewith
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/017,678
;      FILING DATE:  16-MAY-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kitchell, Barbara S.
;      REGISTRATION NUMBER:  33,928
;      REFERENCE/DOCKET NUMBER:  TAMK:193
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  512 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-856-253-6

  Query Match             94.0%;  Score 1054.5;  DB 3;  Length 512;
  Best Local Similarity   64.7%;  Pred. No. 2.1e-87;
  Matches  211;  Conservative    0;  Mismatches    0;  Indels  115;  Gaps    1;

Qy          1 MRGSHHHHHHG---------------------------------------------------- 11
              ||||||||||| 
Db          1 MRGSHHHHHHGSARDISSTNVTDLTVSPSKIEDGGKTTVKMTFDDKNGKIQNGDMIKVAW 60

Qy         12 ------------------------------------------------------------ 11

Db         61 PTSGTVKIEGYSKTVPLTVKGEQVGQAVITPDGATITFNDKVEKLSDVSGFAEFEVQGRN 120

Qy         12 ------SDDKVATITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFLNINNE 65
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LTQTNTSDDKVATITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFLNINNE 180

Qy         66 KSYVSKDITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNT 125
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KSYVSKDITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNT 240

Qy        126 KNTIDVTIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVH 185
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KNTIDVTIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVH 300

Qy        186 NINANAGIEGTVKGELKVLKQDKDTK 211
              ||||||||||||||||||||||||||
Db        301 NINANAGIEGTVKGELKVLKQDKDTK 326


RESULT 3
US-08-447-031A-2
; Sequence 2, Application US/08447031A
; Patent No. 5851794
;  GENERAL INFORMATION:
;    APPLICANT:  GUSS, Bengt
;    APPLICANT:  HOOK, Magnus
;    APPLICANT:  JONSSON, Hans
;    APPLICANT:  LINDBERG, Martin
;    APPLICANT:  PATTI, Joseph
;    APPLICANT:  SIGNAS, Christer
;    APPLICANT:  SWITALSKI, Lech
;    TITLE OF INVENTION:  A COLLAGEN BINDING PROTEIN AS WELL AS
;    TITLE OF INVENTION:  ITS PREPARATION
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/447,031A
;      FILING DATE:  22-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/861,804
;      FILING DATE:  21-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/SE91/00707
;      FILING DATE:  22-OCT-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  SE 9003374-7
;      FILING DATE:  22-OCT-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  012889-006
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1183 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  protein
US-08-447-031A-2

  Query Match             93.4%;  Score 1048;  DB 2;  Length 1183;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-86;
  Matches  200;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          12 SDDKVATITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSK 71
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         144 SDDKVATITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSK 203

Qy          72 DITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDV 131
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         204 DITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDV 263

Qy         132 TIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINANA 191
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         264 TIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINANA 323

Qy         192 GIEGTVKGELKVLKQDKDTK 211
               ||||||||||||||||||||
Db         324 GIEGTVKGELKVLKQDKDTK 343


RESULT 4
US-08-856-253-2
; Sequence 2, Application US/08856253
; Patent No. 6288214
;  GENERAL INFORMATION:
;    APPLICANT:  Hook, Magnus
;    APPLICANT:  Patti, Joseph M.
;    APPLICANT:  House-Pompeo, Karen
;    APPLICANT:  Sthanam, Narayana
;    APPLICANT:  Symersky, Jindrich
;    TITLE OF INVENTION:  COLLAGEN BINDING PROTEIN COMPOSITIONS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  U.S.
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/856,253
;      FILING DATE:  Concurrently Herewith
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/017,678
;      FILING DATE:  16-MAY-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kitchell, Barbara S.
;      REGISTRATION NUMBER:  33,928
;      REFERENCE/DOCKET NUMBER:  TAMK:193
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  159 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-856-253-2

  Query Match             74.0%;  Score 830;  DB 3;  Length 159;
  Best Local Similarity   96.4%;  Pred. No. 8.9e-68;
  Matches  159;  Conservative    0;  Mismatches    0;  Indels    6;  Gaps    1;

Qy           1 MRGSHHHHHHGSDDKVATITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFL 60
               ||||||||||||      ||||||||||||||||||||||||||||||||||||||||||||
Db           1 MRGSHHHHHHGS------ITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFL 54

Qy          61 NINNEKSYVSKDITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKI 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          55 NINNEKSYVSKDITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKI 114

Qy         121 TVDNTKNTIDVTIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQA 165
               |||||||||||||||||||||||||||||||||||||||||||||
Db         115 TVDNTKNTIDVTIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQA 159


RESULT 5
US-09-107-532A-6999
; Sequence 6999, Application US/09107532A
; Patent No. 6583275
;  GENERAL INFORMATION:
;       APPLICANT: Lynn A Doucette-Stamm  and David Bush
;       TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                           ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;       NUMBER OF SEQUENCES: 7310
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;            STREET: 100 Beaver Street
;            CITY: Waltham
;            STATE: Massachusetts
;            COUNTRY: USA
;            ZIP: 02354
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: CD/ROM ISO9660
;            COMPUTER: PC
;            OPERATING SYSTEM: <Unknown>
;            SOFTWARE: ASCII
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/107,532A
;            FILING DATE: 30-Jun-1998
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 60/085,598
;            FILING DATE: 14 May 1998
;            APPLICATION NUMBER: 60/051571
;            FILING DATE: July 2, 1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ariniello, Pamela Deneke
;            REGISTRATION NUMBER: 40,489
;            REFERENCE/DOCKET NUMBER: GTC-012
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (781)893-5007
;            TELEFAX: (781)893-8277
;  INFORMATION FOR SEQ ID NO: 6999:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 731 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       HYPOTHETICAL: YES
;       ORIGINAL SOURCE:
;            ORGANISM: Enterococcus faecium
;       FEATURE:
;            NAME/KEY:  misc_feature
;            LOCATION: (B) LOCATION 1...731
;       SEQUENCE DESCRIPTION: SEQ ID NO: 6999:
US-09-107-532A-6999

  Query Match             43.8%;  Score 491;  DB 4;  Length 731;
  Best Local Similarity   48.5%;  Pred. No. 3.8e-36;
  Matches   95;  Conservative   35;  Mismatches   64;  Indels    2;  Gaps    2;

Qy          18 TITSGNKSTNVTVHKSEAG-TSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSKDITIK 76
               |:|||:|:  | | |  :| :|||||||||||||||| |:||||||||  :|| : : |
Db         162 TVTSGDKTATVNVTKPASGSSSSVFYYKTGDMLPEDTKHIRWFLNINNNGTYVEQPVKIS 221
```

```
Qy          77 DQIQGGQQLDLSTLNIN-VTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIPQ 135
               |:|| ||:|| ||  || : |  | :|: |  | : || :     | | |::|||:
Db         222 DEIQSGQRLDPSTFEINQIHLGEQKVYRGEEGIQQFLQDFPSATFNFSVTDNYIEITIPK 281

Qy         136 GYGSYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINANAGIEG 195
                : :     ::||| | | :|  | |:|:||::|  |  |:|:|||||| ||:|: |: |
Db         282 NFVNLRKIMVSYKTIIENPEQINFENHSEAWFKEFNKPAVDGESFNHTVKNISASGGVNG 341

Qy         196 TVKGELKVLKQDKDTK 211
               ||:||||: |   ||:
Db         342 TVRGELKIFKYINDTE 357


RESULT 6
US-09-134-000C-3745
; Sequence 3745, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;  TITLE OF INVENTION:  ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 032796-032
;  CURRENT APPLICATION NUMBER: US/09/134,000C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/055,778
;  PRIOR FILING DATE: 1997-08-15
;  NUMBER OF SEQ ID NOS: 6812
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3745
;   LENGTH: 540
;   TYPE: PRT
;   ORGANISM: Enterococcus faecalis
US-09-134-000C-3745

  Query Match             15.9%;  Score 178;  DB 4;  Length 540;
  Best Local Similarity   28.3%;  Pred. No. 5.7e-08;
  Matches   54;  Conservative   33;  Mismatches   88;  Indels   16;  Gaps    6;

Qy          18 TITSGNKSTNVTVHKSEAG---TSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSKDIT 74
               | |:  : |   |  :| |       |:|| ||:  | :  ||||||:|  || |::||:
Db         154 TATATQRLTIEGVTNTETGQIERDYPFFYKVGDLAGE-SNQVRWFLNVNLNKSDVTEDIS 212

Qy          75 IKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIP 134
               | |:   ||||:  :   ::      |    :: :||:   |    |  | |  ::
Db         213 IADRQGSGQQLNKESFTFDIVNDKETKY---ISLAEFEQQGYGKIDFV--TDNDFNLRFY 267

Qy         135 QGYGSYNSFSINYKTKIT--NEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINANAG 192
               :   : || : | : ||   : |  | |:    || : ::  | |: :        |
Db         268 RDKARFTSFIVRYTSTITEAGQHQATFENSYDINYQLNNQDATNEKNTSQV-----KNVF 322

Qy         193 IEGTVKGELKV 203
               :||   |   |
Db         323 VEGEASGNQNV 333


RESULT 7
US-08-856-253-7
; Sequence 7, Application US/08856253
; Patent No. 6288214
;  GENERAL INFORMATION:
;    APPLICANT:  Hook, Magnus
;    APPLICANT:  Patti, Joseph M.
;    APPLICANT:  House-Pompeo, Karen
;    APPLICANT:  Sthanam, Narayana
;    APPLICANT:  Symersky, Jindrich
;    TITLE OF INVENTION:  COLLAGEN BINDING PROTEIN COMPOSITIONS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  U.S.
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/856,253
;      FILING DATE:  Concurrently Herewith
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/017,678
;      FILING DATE:  16-MAY-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kitchell, Barbara S.
;      REGISTRATION NUMBER:  33,928
;      REFERENCE/DOCKET NUMBER:  TAMK:193
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  345 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-856-253-7

  Query Match             11.2%;  Score 125.5;  DB 3;  Length 345;
  Best Local Similarity   30.1%;  Pred. No. 0.0017;
  Matches   56;  Conservative   20;  Mismatches   73;  Indels   37;  Gaps   10;

Qy           1 MRGSHHHHHHGS----DDKVATITSGNKSTNV--------TVHKSEAGTSSVFYYKTGDM 48
               ||||||||||||    |   |     |: |||         ||:  :||  : |   |
Db           1 MRGSHHHHHHGSMVAADAPAAGTDITNQLTNVTVGIDSGTTVYPHQAGYVKLNY---GFS 57

Qy          49 LPEDTTHVRWF-------LNINNEKSYVS-KDITIKDQIQGGQQLDLSTLNINVTGTHSN 100
               :|        |       ||:|   |      |   ||:     :|      ||  | ::
Db          58 VPNSAVKGDTFKITVPKELNLNGVTSTAKVPPIMAGDQVLANGVIDSDG---NVIYTFTD 114

Qy         101 YYSGQSAITDFEKA---FPGSKITVDNTKNTIDVTIPQGYGSYNSFSINYKTKITNEQQK 157
               | : :    |  ||    | : |  :| | | :||:  | ||  :   |     :  |:
Db         115 YVNTK----DDVKATLTMP-AYIDPENVKKTGNVTLATGIGSTTA---NKTVLVDYEKYG 166

Qy         158 EFVNNS 163
               :| | |
Db         167 KFYNLS 172


RESULT 8
US-09-071-035-464
; Sequence 464, Application US/09071035
; Patent No. 6448043
;  GENERAL INFORMATION:
;    APPLICANT:  Gil H. Choi
;    TITLE OF INVENTION:  Enterococcus faecalis Polynucleotides and Polypeptides
;    NUMBER OF SEQUENCES:  496
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,035
```

```
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  A. Anders Brookes
;       REGISTRATION NUMBER:  36,373
;       REFERENCE/DOCKET NUMBER:  PB369P2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (301) 309-8504
;       TELEFAX:  (301) 309-8512
;   INFORMATION FOR SEQ ID NO:  464:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  688 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-09-071-035-464

  Query Match             10.2%;  Score 114.5;  DB 4;  Length 688;
  Best Local Similarity   25.5%;  Pred. No. 0.045;
  Matches   54;  Conservative   36;  Mismatches   93;  Indels   29;  Gaps    9;

Qy          12 SDDKVATITSGNKSTNVTVHKS-EAGTSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVS 70
               | |  |: |:   | :||   | : |  ||   | |:   :|  || | : ||  :| |
Db         172 SYDNTASYTNQGSSRDVTGKVSIQHGGESV--KKGGEYHKDDPDHVYWHVMINGAQS-VL 228

Qy          71 KDITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKN-TI 129
                |: | |     | ||  :| |   ||:         :|:     |  : ::  |: |:
Db         229 DDVVITDTPSPNQVLDPESLVI--YGTN---------VTEDGTITPDKSVILEEGKDYTL 277

Qy         130 DVTIPQGYGSYN----------SFSINYKTKITNEQ--QKEFVNNSQAWYQEHGKEEVNG 177
               :||     |              : : |:: :|:      : |:| |     :| | |:|
Db         278 EVTTDNETGQQKIVVKMAHIEAPYYMEYRSLVTSSAAGSTDTVSN-QVSITGNGSEVVHG 336

Qy         178 KSFNHTVHNINANAGIEGTVKGELKVLKQDKD 209
                   | :|: : |     ||:::: |   |
Db         337 DDNGDVVVDIDHSGGHATGTKGKIQLKKTAMD 368


RESULT 9
US-09-071-035-458
; Sequence 458, Application US/09071035
; Patent No. 6448043
;  GENERAL INFORMATION:
;    APPLICANT:  Gil H. Choi
;    TITLE OF INVENTION:  Enterococcus faecalis Polynucleotides and Polypeptides
;    NUMBER OF SEQUENCES:  496
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,035
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  A. Anders Brookes
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB369P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  458:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2032 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-071-035-458

  Query Match             10.2%;  Score 114.5;  DB 4;  Length 2032;
  Best Local Similarity   25.5%;  Pred. No. 0.2;
  Matches   54;  Conservative   36;  Mismatches   93;  Indels   29;  Gaps    9;

Qy          12 SDDKVATITSGNKSTNVTVHKS-EAGTSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVS 70
               | |  |: |:   | :||   | : |  ||   | |:   :|  || | : ||  :| |
Db         908 SYDNTASYTNQGSSRDVTGKVSIQHGGESV--KKGGEYHKDDPDHVYWHVMINGAQS-VL 964

Qy          71 KDITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKN-TI 129
                |: | |     | ||  :| |   ||:         :|:     |  : ::  |: |:
Db         965 DDVVITDTPSPNQVLDPESLVI--YGTN---------VTEDGTITPDKSVILEEGKDYTL 1013

Qy         130 DVTIPQGYGSYN----------SFSINYKTKITNEQ--QKEFVNNSQAWYQEHGKEEVNG 177
               :||     |              : : |:: :|:      : |:| |     :| | |:|
Db        1014 EVTTDNETGQQKIVVKMAHIEAPYYMEYRSLVTSSAAGSTDTVSN-QVSITGNGSEVVHG 1072

Qy         178 KSFNHTVHNINANAGIEGTVKGELKVLKQDKD 209
                   | :|: : |     ||:::: |   |
Db        1073 DDNGDVVVDIDHSGGHATGTKGKIQLKKTAMD 1104


RESULT 10
US-09-071-035-462
; Sequence 462, Application US/09071035
; Patent No. 6448043
;  GENERAL INFORMATION:
;    APPLICANT:  Gil H. Choi
;    TITLE OF INVENTION:  Enterococcus faecalis Polynucleotides and Polypeptides
;    NUMBER OF SEQUENCES:  496
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,035
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  A. Anders Brookes
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB369P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  462:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2032 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  protein
US-09-071-035-462

  Query Match             10.2%;  Score 114.5;  DB 4;  Length 2032;
  Best Local Similarity   25.5%;  Pred. No. 0.2;
  Matches   54;  Conservative   36;  Mismatches   93;  Indels   29;  Gaps    9;

Qy         12 SDDKVATITSGNKSTNVTVHKS-EAGTSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVS 70
Db        908 SYDNTASYTNQGSSRDVTGKVSIQHGGESV--KKGGEYHKDDPDHVYWHVMINGAQS-VL 964

Qy         71 KDITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKN-TI 129
Db        965 DDVVITDTPSPNQVLDPESLVI--YGTN---------VTEDGTITPDKSVILEEGKDYTL 1013

Qy        130 DVTIPQGYGSYN----------SFSINYKTKITNEQ--QKEFVNNSQAWYQEHGKEEVNG 177
Db       1014 EVTTDNETGQQKIVVKMAHIEAPYYMEYRSLVTSSAAGSTDTVSN-QVSITGNGSEVVHG 1072

Qy        178 KSFNHTVHNINANAGIEGTVKGELKVLKQDKD 209
Db       1073 DDNGDVVVDIDHSGGHATGTKGKIQLKKTAMD 1104


RESULT 11
US-09-071-035-466
; Sequence 466, Application US/09071035
; Patent No. 6448043
;  GENERAL INFORMATION:
;    APPLICANT:  Gil H. Choi
;    TITLE OF INVENTION:  Enterococcus faecalis Polynucleotides and Polypeptides
;    NUMBER OF SEQUENCES:  496
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,035
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  A. Anders Brookes
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB369P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  466:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2032 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-071-035-466

  Query Match             10.2%;  Score 114.5;  DB 4;  Length 2032;
  Best Local Similarity   25.5%;  Pred. No. 0.2;
  Matches   54;  Conservative   36;  Mismatches   93;  Indels   29;  Gaps    9;

Qy         12 SDDKVATITSGNKSTNVTVHKS-EAGTSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVS 70
Db        908 SYDNTASYTNQGSSRDVTGKVSIQHGGESV--KKGGEYHKDDPDHVYWHVMINGAQS-VL 964

Qy         71 KDITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKN-TI 129
Db        965 DDVVITDTPSPNQVLDPESLVI--YGTN---------VTEDGTITPDKSVILEEGKDYTL 1013

Qy        130 DVTIPQGYGSYN----------SFSINYKTKITNEQ--QKEFVNNSQAWYQEHGKEEVNG 177
Db       1014 EVTTDNETGQQKIVVKMAHIEAPYYMEYRSLVTSSAAGSTDTVSN-QVSITGNGSEVVHG 1072

Qy        178 KSFNHTVHNINANAGIEGTVKGELKVLKQDKD 209
Db       1073 DDNGDVVVDIDHSGGHATGTKGKIQLKKTAMD 1104


RESULT 12
US-09-134-000C-6612
; Sequence 6612, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;  TITLE OF INVENTION:  ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 032796-032
;  CURRENT APPLICATION NUMBER: US/09/134,000C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/055,778
;  PRIOR FILING DATE: 1997-08-15
;  NUMBER OF SEQ ID NOS: 6812
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6612
;   LENGTH: 2054
;   TYPE: PRT
;   ORGANISM: Enterococcus faecalis
US-09-134-000C-6612

  Query Match             10.2%;  Score 114.5;  DB 4;  Length 2054;
  Best Local Similarity   25.5%;  Pred. No. 0.21;
  Matches   54;  Conservative   36;  Mismatches   93;  Indels   29;  Gaps    9;

Qy         12 SDDKVATITSGNKSTNVTVHKS-EAGTSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVS 70
Db        930 SYDNTASYTNQGSSRDVTGKVSIQHGGESV--KKGGEYHKDDPDHVYWHVMINGAQS-VL 986

Qy         71 KDITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKN-TI 129
Db        987 DDVVITDTPSPNQVLDPESLVI--YGTN---------VTEDGTITPDKSVILEEGKDYTL 1035

Qy        130 DVTIPQGYGSYN----------SFSINYKTKITNEQ--QKEFVNNSQAWYQEHGKEEVNG 177
Db       1036 EVTTDNETGQQKIVVKMAHIEAPYYMEYRSLVTSSAAGSTDTVSN-QVSITGNGSEVVHG 1094

Qy        178 KSFNHTVHNINANAGIEGTVKGELKVLKQDKD 209
Db       1095 DDNGDVVVDIDHSGGHATGTKGKIQLKKTAMD 1126


RESULT 13
US-09-023-082A-30
; Sequence 30, Application US/09023082A
; Patent No. 6077692
;  GENERAL INFORMATION:
;    APPLICANT:  RUBEN, STEVEN M.
;    APPLICANT:  JIMENEZ, PABLO
;    APPLICANT:  DUAN, D. ROXANNE
;    APPLICANT:  RAMPY, MARK A.
;    APPLICANT:  MENDRICK, DONNA
;    APPLICANT:  ZHANG, JUN
;    APPLICANT:  NI, JIAN
;    APPLICANT:  MOORE, PAUL A.
;    APPLICANT:  COLEMAN, TIMOTHY A.
;    APPLICANT:  GRUBER, JOACHIM R.
;    APPLICANT:  DILLON, PATRICK J.
```

```
;    APPLICANT:  GENTZ, REINER L.
;    TITLE OF INVENTION:  KERATINOCYTE GROWTH FACTOR-2
;    NUMBER OF SEQUENCES:  148
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;      STREET:  1100 NEW YORK AVE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/023,082A
;      FILING DATE:  13-FEB-1998
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/01790
;      FILING DATE:  14-FEB-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/461,195
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/023,852
;      FILING DATE:  13-AUG-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/039,045
;      FILING DATE:  28-FEB-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/862,432
;      FILING DATE:  23-MAY-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/910,875
;      FILING DATE:  13-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/055,561
;      FILING DATE:  13-AUG-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEFFFE, ERIC K.
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0360008/EKS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  184 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-023-082A-30

  Query Match             10.1%;  Score 113.5;  DB 3;  Length 184;
  Best Local Similarity   23.4%;  Pred. No. 0.0088;
  Matches   45;  Conservative   35;  Mismatches   65;  Indels   47;  Gaps    9;

Qy          1 MRGSHHHHHHGSDDKVATITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFL 60
              ||||||||||||   :       ::|| :     : :|:  : :: : |  |   |||
Db          1 MRGSHHHHHHGSCQALGQDMVSPEATNSSSSSFSSPSSAGRHVRSYNHLQGD---VRWRK 57

Qy         61 NINNEKSYVSKDITIKDQIQGGQQ-------LDLSTLNINVTGT---HSNYYSGQSAITD 110
                 :  | ::   |    :: | ::       |:::: | |       :||||    |:
Db         58 LFSFTKYFLK--IEKNGKVSGTKKENCPYSILEITSVEIGVVAVKAINSNYY---LAMNK 112

Qy        111 FEKAFPGSKITVD-NTKNTIDVTIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQAWYQE 169
                | :   :   |   |  |:     || :| ||:                    :|
Db        113 KGKLYGSKEFNNDCKLKERIE---ENGYNTYASFN--------------------WQH 147

Qy        170 HGKE---EVNGK 178
              :|::    :|||
Db        148 NGRQMYVALNGK 159


RESULT 14
US-09-248-998-30
; Sequence 30, Application US/09248998
; Patent No. 6599879
; GENERAL INFORMATION:
;  APPLICANT: Jimenez, Pablo
;  APPLICANT: Rampy, Mark A.
;  APPLICANT: Mendrick, Donna
;  APPLICANT: Russell, Deborah
;  APPLICANT: Louie, Arthur
;  TITLE OF INVENTION: Therapeutic Uses of Keratinocyte Growth Factor-2
;  FILE REFERENCE: 1488.1060002
;  CURRENT APPLICATION NUMBER: US/09/248,998
;  CURRENT FILING DATE: 1999-02-12
;  EARLIER APPLICATION NUMBER: US 60/114,387
;  EARLIER FILING DATE: 30-DEC-1998
;  EARLIER APPLICATION NUMBER: US 60/074,585
;  EARLIER FILING DATE: 13-FEB-1998
;  NUMBER OF SEQ ID NOS: 148
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
;   LENGTH: 184
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-248-998-30

  Query Match             10.1%;  Score 113.5;  DB 4;  Length 184;
  Best Local Similarity   23.4%;  Pred. No. 0.0088;
  Matches   45;  Conservative   35;  Mismatches   65;  Indels   47;  Gaps    9;

Qy          1 MRGSHHHHHHGSDDKVATITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFL 60
              ||||||||||||   :       ::|| :     : :|:  : :: : |  |   |||
Db          1 MRGSHHHHHHGSCQALGQDMVSPEATNSSSSSFSSPSSAGRHVRSYNHLQGD---VRWRK 57

Qy         61 NINNEKSYVSKDITIKDQIQGGQQ-------LDLSTLNINVTGT---HSNYYSGQSAITD 110
                 :  | ::   |    :: | ::       |:::: | |       :||||    |:
Db         58 LFSFTKYFLK--IEKNGKVSGTKKENCPYSILEITSVEIGVVAVKAINSNYY---LAMNK 112

Qy        111 FEKAFPGSKITVD-NTKNTIDVTIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQAWYQE 169
                | :   :   |   |  |:     || :| ||:                    :|
Db        113 KGKLYGSKEFNNDCKLKERIE---ENGYNTYASFN--------------------WQH 147

Qy        170 HGKE---EVNGK 178
              :|::    :|||
Db        148 NGRQMYVALNGK 159


RESULT 15
US-08-728-470-9
; Sequence 9, Application US/08728470
; Patent No. 5928651
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5928651-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/728,470
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-633
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1338 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-728-470-9

  Query Match             9.8%;  Score 110.5;  DB 2;  Length 1338;
  Best Local Similarity   25.5%;  Pred. No. 0.26;
  Matches   61;  Conservative   28;  Mismatches   89;  Indels   61;  Gaps   10;

Qy         12 SDDKVATITSGNKST-----NVTVHKSEA-----GTSSVFYYKTGDMLPEDTTHVRWFLN 61
              ||    ||  || |      |:   ||:|     |  |    | |:           |
Db        694 SDTTYRTIIKGNISNKSGDLNIIDKKSDAEIQIGGNIS---QKEGN-----------LT 738

Qy         62 INNEKSYVSKDITIKDQIQGGQQLDLSTLNINVT-GTHSNYYSGQSAITDFEKAFPGSKI 120
              |:::|  ::  ||||  ::||:       | |:|  |     :|   |: | ||    :|
Db        739 ISSDKVNITNQITIKAGVEGGRSDSSEAENANLTIQTKELKLAGDLNISGFNKA----EI 794

Qy        121 TVDNTKNTIDVTIPQGYGSY-------------------NSFSINYKTKITNEQQKEFV 160
              |    ||  |:||    |                   :: ::| : | :|
Db        795 T---AKNGSDLTIGNASGGNADAKKVTFDKVKDSKISTDGHNVTLNSEVKTSNGSSNAGN 851

Qy        161 NNSQAWYQEHGKEEVNGKSFNHTVHNINANAG----IEGT----VKGELKVLKQDKDTK 211
              :||          ||    :|   ||:| ||     |||      | ::|  |:   |
Db        852 DNSTGLTISAKDVTVNNNVTSHKTINISAAAGNVTTKEGTTINATTGSVEVTAQNGTIK 910


Search completed: July 27, 2004, 10:10:29
Job time : 20 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 27, 2004, 10:06:36 ; Search time 16 Seconds
(without alignments)
1268.526 Million cell updates/sec

Title: US-09-813-820-4
Perfect score: 1122
Sequence: 1 MRGSHHHHHHGSDDKVATIT.....GIEGTVKGELKVLKQDKDTK 211

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	911	81.2	1185	2	A42404	collagen adhesin -
2	194	17.3	37	2	A48620	adhesin - Staphylo
3	131	11.7	1092	2	S42798	fibronectin-bindin
4	119	10.6	336	2	D86710	collagen adhesin [
5	101.5	9.0	542	2	A98021	hypothetical prote
6	100	8.9	1530	2	AH1396	peptidoglycan anch
7	100	8.9	1536	2	A43855	high-molecular-wei
8	99.5	8.9	266	2	D97203	hypothetical prote
9	99.5	8.9	463	2	AG1542	wall associated pr
10	99.5	8.9	592	2	A81888	probable surface f
11	98.5	8.8	4919	2	T31105	hypothetical prote
12	98	8.7	1117	2	S33851	fibronectin-bindin
13	98	8.7	1248	2	C89874	autolysin [importe
14	96.5	8.6	848	2	C70203	DNA topoisomerase
15	96	8.6	392	2	A53580	neurexin III beta
16	96	8.6	704	1	S39004	finger protein MSN
17	95	8.5	359	2	E86706	hypothetical prote
18	95	8.5	666	2	A42296	lysozyme 2 (EC 3.2
19	95	8.5	2020	2	C48399	ABC-type transport
20	94.5	8.4	188	2	E90065	hypothetical prote
21	94.5	8.4	2902	2	C71953	toxin-like outer m
22	94.5	8.4	3890	2	C89921	hypothetical prote
23	94	8.4	2380	2	E71604	hypothetical prote
24	93.5	8.3	532	2	T06029	hypothetical prote
25	93.5	8.3	1297	2	S39791	neurotoxin - Clost
26	93.5	8.3	2340	2	B71704	cell surface antig
27	92.5	8.2	681	2	AC1458	surface anchored p
28	92.5	8.2	711	2	S73898	DNA topoisomerase
29	92.5	8.2	1943	2	B64596	toxin-like outer m
30	92.5	8.2	2910	2	T28156	DNA-directed RNA p
31	92	8.2	1037	2	T13350	transcription fact
32	92	8.2	1983	2	G86643	hypothetical prote
33	92	8.2	2500	2	G71609	hypothetical prote
34	91.5	8.2	261	2	S16619	opacity protein op
35	91.5	8.2	622	2	AF0169	probable exported
36	91.5	8.2	2150	2	S71629	sensory transducti
37	91.5	8.2	2271	2	F90073	hypothetical prote
38	91	8.1	452	2	AE1710	hypothetical prote
39	91	8.1	463	2	A44808	cellulase (EC 3.2.
40	91	8.1	471	2	H97200	ferrichrome-bindin
41	91	8.1	578	2	B90578	lipoprotein a [imp
42	91	8.1	953	2	C89824	hypothetical prote
43	91	8.1	2269	2	T28677	rhoptry protein -
44	90.5	8.1	784	2	JH0101	apolipoprotein B-1
45	90	8.0	669	2	S14535	asparagine-rich pr

ALIGNMENTS

RESULT 1
A42404
collagen adhesin - Staphylococcus aureus
C;Species: Staphylococcus aureus
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 10-Nov-1995
C;Accession: A42404; S27665
R;Patti, J.M.; Jonsson, H.; Guss, B.; Switalski, L.M.; Wiberg, K.; Lindberg, M.; Hook, M
J. Biol. Chem. 267, 4766-4772, 1992
A;Title: Molecular characterization and expression of a gene encoding a Staphylococcus a
A;Reference number: A42404; MUID:92165839; PMID:1311320
A;Contents: FDA 574
A;Accession: A42404
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1185 <PAT>
A;Cross-references: EMBL:M81736
A;Note: sequence extracted from NCBI backbone (NCBIP:83982)

Query Match 81.2%; Score 911; DB 2; Length 1185;
Best Local Similarity 90.1%; Pred. No. 7e-58;
Matches 182; Conservative 2; Mismatches 16; Indels 2; Gaps 2;

```
Qy          12 SDDKVATITSGNKSTNVTVHKSEAGTSSVFYY-KTGDM-LPEDTTHVRWFLNINNEKSYV 69
               ||||||||||||||||||            ||   |:| :   ||||||||||||||||||||
Db         144 SDDKVATITSGNKSTNVIGWIKVKREPVVFLINKSGKICYQEDTTHVRWFLNINNEKSYV 203

Qy          70 SKDITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTI 129
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         204 SKDITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTI 263

Qy         130 DVTIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINA 189
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         264 DVTIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINA 323

Qy         190 NAGIEGTVKGELKVLKQDKDTK 211
               ||||||||||||||||||||||
Db         324 NAGIEGTVKGELKVLKQDKDTK 345
```

RESULT 2
A48620
adhesin - Staphylococcus aureus (fragment)
C;Species: Staphylococcus aureus
C;Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 03-Mar-1995
C;Accession: A48620
R;Patti, J.M.; Boles, J.O.; Hook, M.
Biochemistry 32, 11428-11435, 1993
A;Title: Identification and biochemical characterization of the ligand binding domain of
A;Reference number: A48620; MUID:94032261; PMID:8218209
A;Contents: FDA 574
A;Accession: A48620
A;Status: preliminary; not compared with conceptual translation

```
A;Molecule type: nucleic acid; protein
A;Residues: 1-37 <PAT>
A;Note: sequence extracted from NCBI backbone (NCBIP:138726)

  Query Match             17.3%;  Score 194;  DB 2;  Length 37;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-08;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         19 ITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTH 55
              |||||||||||||||||||||||||||||||||||||
Db          1 ITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTH 37


RESULT 3
S42798
fibronectin-binding protein - Streptococcus "equisimilis"
C;Species: Streptococcus "equisimilis"
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 15-Oct-1999
C;Accession: S42798
R;Signaes, C.
submitted to the EMBL Data Library, December 1993
A;Reference number: S42798
A;Accession: S42798
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1092 <SIG>
A;Cross-references: EMBL:Z29088; NID:g436135; PIDN:CAA82330.1; PID:g436136
C;Keywords: fibronectin binding

  Query Match             11.7%;  Score 131;  DB 2;  Length 1092;
  Best Local Similarity   26.5%;  Pred. No. 0.13;
  Matches   58;  Conservative   31;  Mismatches   88;  Indels   42;  Gaps   13;

Qy         19 ITSGNKSTNVTVHKSEAGTSS-VFYYKTGDMLPEDTTH---VRWFLNIN-NEKSYVSKDI 73
              : || |   | : |   ||::   |||:||:   || :   |||    |   ||
Db        152 VISGTKHAEVVIQKERGGTATDNIYYKSGDI--PDTQNEDSVRWSFTFNAARKSTNGVGF 209

Qy         74 TIKDQIQGGQQLDL-----STLNINVTGTHSNYYSGQSAITDFE-----KAFPGSKITVD 123
              : | :     |:          || ||     : | :       |     | |:
Db        210 LVTDTLDSTMTWDVEKNLREKYAINFTGG----WIGDELLKSKNGWLSLKEAENYGIKVE 265

Qy        124 NTKNTIDVTIPQ----GYGSY-----NSFSINYKTKITNEQQKEFVNNSQAWY-QEHGKE 173
               :  |: : ||:      |||     :   |:   ||    :|| : ||   | :   |
Db        266 FSGQTVSINIPEKIQYDNGSYQELNESEMQIHLYAKI----KKEVLENSSIEYVRNESKV 321

Qy        174 EVNGKSF----NHT---VHNINANAGIEGTVKGELKVLK 205
              || |:::    | |   |  :      :|||:||:::||
Db        322 EVRGENWPIDPNSTSAYVQILRQGGWAKGTVRGEVRILK 360


RESULT 4
D86710
collagen adhesin [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C;Accession: D86710
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: D86710
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-336 <STO>
A;Cross-references: GB:AE005176; PID:g12723592; PIDN:AAK04782.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: icaC

  Query Match             10.6%;  Score 119;  DB 2;  Length 336;
  Best Local Similarity   25.2%;  Pred. No. 0.22;
  Matches   52;  Conservative   28;  Mismatches   68;  Indels   58;  Gaps   11;

Qy         24 KSTNVTVHKSEAGTSSV------------FYYKTGDMLPEDTTHVRWFLNINNEKSYVSK 71
              : | | |   : ||: |              || | :   |:| : | : ||  |  : |
Db         71 EDTRVVVDWKDGGTTEVNIGPATGPDKDEVLYKWGWVDENDSTLIHWQVRINYAKENIQK 130

Qy         72 DITIKDQIQGGQQL---DLSTLNINVTGTHSNY----YSGQSAITDFEKAFPGSKITVDN 124
               |   | | | |     :|  |: :    ||     |  |::|    |    |  : : :
Db        131 AI-YTDIIGGNQNLVSGSISVANVTYSSDGENYNVDSYYPQASI--LENGVNGFTVNLGD 187

Qy        125 TKNTIDVTIPQGYGSYNSFSINYKTKITN---EQQKEFVNNSQAWYQEHGKEEVNGKSFN 181
                |||               |:| |: |:    ||          |:  |  |: |::
Db        188 ISNTI--------------IIDYSTRATDGGLSQQ----------YENRG--ELTGENIE 221

Qy        182 ------HTVHN-INANAGIEGTVKGE 200
                    || :|  | || :: :: ||
Db        222 KQVVEVHTPNNGGNGNASMKLSISGE 247


RESULT 5
A98021
hypothetical protein appA [imported] - Streptococcus pneumoniae (strain R6)
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C;Accession: A98021
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: A98021
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-542 <KUR>
A;Cross-references: GB:AE007317; PIDN:AAK99997.1; PID:g15458827; GSPDB:GN00174
C;Genetics:
A;Gene: appA

  Query Match             9.0%;  Score 101.5;  DB 2;  Length 542;
  Best Local Similarity   25.3%;  Pred. No. 7.2;
  Matches   49;  Conservative   28;  Mismatches   62;  Indels   55;  Gaps   10;

Qy         20 TSGNKSTNVTVHKSEAGTSSVFYYK-----TGDMLP------------------------ 50
              :||:| ::     ||:|   |||       ||| :|
Db         25 SSGSKKSDF----SESGKEKVFYLSGFLQGTGDRIPYVWQNGDGLVPYLLYRTLLMADSK 80

Qy         51 -EDTTH--VRWFLNINNEKSYVSKDITIKDQIQ--GGQQLDLSTLNINV---------TG 96
               | :|    : :   ::||:|   : |:|| ::    |: |  | :  :|          |
Db         81 YEKSTPDLAKDWKKSDDEKTY---EFTLKDGLKWSDGEALTASDVEFSVKLALKASVING 137

Qy         97 THSNYYSGQSAITDF--EKAFPGSKITVDNTKNTIDVTIPQG--YGSYNSFSINYKTKIT 152
                 : ::      ||  :||   | :|||  | |  :| | |        | |  | |:
Db        138 IFPSTFTKIEGAKDFKDDKADKISGLTVDGNKVTFKLTAPVGDFEQMMTQFFILPKHKLE 197

Qy        153 NEQQKEFVNNSQAW 166
              :|   | :|||  |
Db        198 SENPLE-LNNSSFW 210


RESULT 6
AH1396
peptidoglycan anchored protein (LPXTG motif) [imported] - Listeria monocytogenes (strain
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AH1396
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
```

A;Accession: AH1396
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1530 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAD00654.1; PID:g16412064; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo2576

Query Match 8.9%; Score 100; DB 2; Length 1530;
Best Local Similarity 22.4%; Pred. No. 34;
Matches 49; Conservative 33; Mismatches 73; Indels 64; Gaps 10;

```
Qy         17 ATITSG-NKSTNVTVHKSEAGTSS------------------VFYYKTGDMLPEDTTHV 56
Db        555 STHTSGDNEYTNNATIKTDTAEASDSDTKWIDKVIDADGYKNGVFNYKTGE--------I 606

Qy         57 RWFLNINNEKSYVSKDITIKDQIQGGQ----------QLDLSTL----------NINVTG 96
Db        607 EWKLIFNDSSKLISKP-TIEDSLNSGQTFIQDSIEIHKIDLSATPQVGELIPPENYDVTF 665

Qy         97 THSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIPQGYGSYNSFSI---------NY 147
Db        666 TKKENGNEQMLIT-FKKPLI-HPVEVTYKTKPVGITKP----LYKNKAVISDGEEVLADY 719

Qy        148 KTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHN 186
Db        720 EAEVIDDNANKYVNKSGEQVGDNIDWEIYANQSGSTVSN 758
```

RESULT 7
A43855
high-molecular-weight surface-exposed protein HMW1 - Haemophilus influenzae
C;Species: Haemophilus influenzae
C;Date: 10-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
C;Accession: A43855
R;Barenkamp, S.J.; Leininger, E.
Infect. Immun. 60, 1302-1313, 1992
A;Title: Cloning, expression, and DNA sequence analysis of genes encoding nontypeable Ha
detella pertussis.
A;Reference number: A43855; MUID:92192797; PMID:1548058
A;Accession: A43855
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1536 <BAR>
A;Cross-references: GB:U08876; GB:M84616; NID:g475770; PIDN:AAA20527.1; PID:g475771
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:89235, NCBIP:89239)

Query Match 8.9%; Score 100; DB 2; Length 1536;
Best Local Similarity 25.2%; Pred. No. 34;
Matches 51; Conservative 33; Mismatches 72; Indels 46; Gaps 9;

```
Qy         13 DDKVATITSGNKSTNVTVHKSEAGTSSVFYYKTGDM------------LPEDTTHVRWFL 60
Db        943 NSKNLSITTNSSSTYRTI------ISGNITNKNGDLNITNEGSDTEMQIGGDVSQKEGNL 996

Qy         61 NINNEKSYVSKDITIKDQIQGGQQLDLSTLNINVT-GTHSNYYSGQSAITDFEK----AF 115
Db        997 TISSDKINITKQITIKAGVDGENSDSDATNNANLTIKTKELKLTQDLNISGFNKAEITAK 1056

Qy        116 PGSKITVDNTKNTIDVTIPQGYGSYNSFSINYK----TKITNEQQKEFVNNSQAWYQEHG 171
Db       1057 DGSDLTIGNT-NSADGT--------NAKKVTFNQVKDSKISADGHKVTL---------HS 1098

Qy        172 KEEVNGKSFNHTVHNINANAGI 193
Db       1099 KVETSG-SNNNTEDSSDNNAGL 1119
```

RESULT 8
D97203
hypothetical protein CAC2461 [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: D97203
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: D97203
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-266 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK80415.1; PID:g15025480; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC2461

Query Match 8.9%; Score 99.5; DB 2; Length 266;
Best Local Similarity 22.2%; Pred. No. 4.1;
Matches 50; Conservative 39; Mismatches 83; Indels 53; Gaps 11;

```
Qy          9 HHGSDD-------KVATITSGNKSTNVTVHKSEAGTSSVFYY---------------KTG 46
Db         50 YYNSNDGKEVERSKIITHPTNNKSKPFIVERVKIGTDTKITYEDDYSCRLIEPMVDKKVK 109

Qy         47 DMLPEDTTHVRW--------FLNINNEKSYVSKDITIKDQIQGGQQLDLSTLNINVTGTH 98
Db        110 NILGNDSSNIKCASYVSAEDYQNLDDYTSKIDYDAFIKDQ---GNN---STINIIAVVNH 163

Qy         99 SNYYSGQSAITDFEKAFPGSKITVD----NTKNTIDVTI----PQGYGSYNSFSINYKTK 150
Db        164 GDNFDKN---TEAQKVYSFMKKIVDTKLVKTHMHIEVSFYFVKSQYYNKTNPQAIAYRYN 220

Qy        151 ITNEQQKEFVN---NSQAWYQEHG--KEEVNGKSFNHTVHNINAN 190
Db        221 -RSMMDSEWLNKNENTTKNYVDHSFTLRSEDGLSYNQDINEIMKN 264
```

RESULT 9
AG1542
wall associated protein precursor (LPXTG motif) homolog lin0879 [imported] - Listeria ir
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AG1542
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AG1542
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-463 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC96111.1; PID:g16413329; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin0879

Query Match 8.9%; Score 99.5; DB 2; Length 463;
Best Local Similarity 24.3%; Pred. No. 8.3;
Matches 50; Conservative 33; Mismatches 78; Indels 45; Gaps 12;

```
Qy         19 ITSGNKSTNVTVHKSEAG---TSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSKDITI 75
Db        153 VNNTTQTINTYISKVTTGGGTGSPTIVFKQGRMDDKDTSIIHWTVTLNNALTPIDNAI-Y 211

Qy         76 KDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIPQ 135
Db        212 TDTLGTGQNL-LGNATIKYRDANKKVLS-----TNVQ------PIALDSNRN-FELTI-- 256
```

```
Qy        136 GYGSYNSFS--INYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINAN--- 190
Db        257 --GTLNNQSVVITYDTKITT-KQKSYTN----------KATLSGDNLDAVSRNATVNDYG 303

Qy        191 AGIEGT--------VKGELKVLKQDK 208
Db        304 SGGQGTGTPPPTPPVKEEPPFIPADK 329
```

RESULT 10
A81888
probable surface fibril protein NMA1200 [imported] - Neisseria meningitidis (strain Z249
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C;Accession: A81888
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria menigitidis Z2491.
A;Reference number: A81775; MUID:20222556; PMID:10761919
A;Accession: A81888
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-592 <PAR>
A;Cross-references: GB:AL162755; GB:AL157959; NID:g7379742; PIDN:CAB84461.1; PID:g737989
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: NMA1200

  Query Match 8.9%; Score 99.5; DB 2; Length 592;
  Best Local Similarity 21.6%; Pred. No. 11;
  Matches 43; Conservative 29; Mismatches 90; Indels 37; Gaps 6;

```
Qy         11 GSDDKVATITSGNKSTN-----------VTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWF 59
Db        147 GANGKKVNIISDTKGLNFAKETAGTNGDTTVHLNGIGST-----LTDTLAGSSASHV--- 198

Qy         60 LNINNNEKSYVSKDITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSK 119
Db        199 -DAGNQSTHYTRAASIKDVLNAGW---------NIKGVKTGSTTGQSENVDFVRTYDTVE 248

Qy        120 ITVDNTKNTIDVTIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKS 179
Db        249 FLSADTKTTTVNVESKDNGKRTEVKIGAKTSVIKEKDGKLVTGK-------GKGE-NGSS 300

Qy        180 FNHTVHNINANAGIEGTVK 198
Db        301 TDEGEGLVTAKEVIDAVNK 319
```

RESULT 11
T31105
hypothetical protein 2 - Haemophilus ducreyi
C;Species: Haemophilus ducreyi
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C;Accession: T31105
R;Ward, C.K.; Lumbley, S.R.; Latimer, J.L.; Cope, L.D.; Hansen, E.J.
J. Bacteriol. 180, 6013-6022, 1998
A;Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.
A;Reference number: Z20984; MUID:99030326; PMID:9811662
A;Accession: T31105
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4919 <WAR>
A;Cross-references: EMBL:AF057696; NID:g3929021; PID:g3929023; PIDN:AAC79761.1
C;Genetics:
A;Gene: lspA2

  Query Match 8.8%; Score 98.5; DB 2; Length 4919;
  Best Local Similarity 24.8%; Pred. No. 1.9e+02;
  Matches 53; Conservative 34; Mismatches 70; Indels 57; Gaps 12;

```
Qy         15 KVATITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSKDIT 74
Db       1104 RIGKIYAGR---DLTFNKSNAGGKS-------EIINRGT------INVKNKLSYDS-DVS 1146

Qy         75 IKDQIQGGQQLDLST----------LNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDN 124
Db       1147 FENNMQ-SQKVDLYTKIFEAKSDIELTFKTNGTHPVYLNFKS--NNNEKKYRNS----EN 1199

Qy        125 TKN----------TIDVTIPQGYGSYNS-FSINYKTKITNEQQKEFVNNSQAWYQEHGKE 173
Db       1200 TKNFKSIGDLINEALSDSAPEAIEAYYSGSSSNYINPVSYLAALGNANNSS--------- 1250

Qy        174 EVNGKSFNHTVHNINANAGIEGTVKGE-LKVLKQ 206
Db       1251 --NPHYLNTALKHILGNGWQDDLKKQENIKVLKQ 1282
```

RESULT 12
S33851
fibronectin-binding protein precursor - Streptococcus dysgalactiae
C;Species: Streptococcus dysgalactiae
C;Date: 31-Dec-1993 #sequence_revision 02-Aug-1994 #text_change 15-Oct-1999
C;Accession: S33851; S32631
R;Lindgren, P.E.; McGavin, M.J.; Signaes, C.; Guss, B.; Gurusiddappa, S.; Hoeoek, M.; Li
Eur. J. Biochem. 214, 819-827, 1993
A;Title: Two different genes coding for fibronectin-binding proteins from Streptococcus
A;Reference number: S33850; MUID:93307299; PMID:8319691
A;Accession: S33851
A;Molecule type: DNA
A;Residues: 1-1117 <LIN>
A;Cross-references: EMBL:Z22151; NID:g288970; PIDN:CAA80122.1; PID:g581643
A;Note: the authors translated the initiation codon TTG for residue 1 as Leu
C;Genetics:
A;Gene: fnbB
A;Start codon: TTG
C;Keywords: fibronectin binding
F;1-27/Domain: signal sequence #status predicted <SIG>
F;28-1117/Product: fibronectin-binding protein #status predicted <MAT>

  Query Match 8.7%; Score 98; DB 2; Length 1117;
  Best Local Similarity 23.3%; Pred. No. 32;
  Matches 51; Conservative 35; Mismatches 85; Indels 48; Gaps 11;

```
Qy         27 NVTVHKSEAGTSSVFYYKTGDMLPEDTTH--VRWFLNINNEKSYVSKDITIK----DQIQ 80
Db        157 NINYNAGGSRTDSV-YAKSGVLDDFDSTGRIVNWYFNFNAALNPI-KDEDVEFMTTDNLP 214

Qy         81 GGQQLDLSTLN-----INVTGTH-SNYYSGQSAITDFEKA---------FPGSKITVDNTK 126
Db        215 QTMTWDIEAVKKDPYVVQIRGGYLGTVFSKNGGWIDIEEATKLGIEIIFSGQKLT----- 269

Qy        127 NTIDVTIPQGYGSYN------------SFSINYKTKIT----NEQQKEFVNNSQAWYQEH 170
Db        270 ----IKIPK-WTNYNGLIPYVKPLNQTSISVKLTAKITEKTMRDQTIEYVENDSIPKIEG 324

Qy        171 GKEEVNGKSFNHTVHNINANAGIEGTVKGELKVLKQDKD 209
Db        325 IEWNLDKNQFGAKVRIQRSGSWILGVRPGDLKVLKFERD 363
```

RESULT 13
C89874
autolysin [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: C89874
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Stapylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: C89874
A;Status: preliminary

```
A;Molecule type: DNA
A;Residues: 1-1248 <KUR>
A;Cross-references: GB:BA000018; PID:g13700854; PIDN:BAB42150.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: atl

  Query Match             8.7%;  Score 98;  DB 2;  Length 1248;
  Best Local Similarity   21.6%;  Pred. No. 37;
  Matches   61;  Conservative   39;  Mismatches   90;  Indels   92;  Gaps   16;

Qy         15 KVATITSGNKSTNVTVHKSEAGTSSVF----YYKTGD--------MLPEDTTH---VRWF 59
              |:| :   |     :|::  |   : :    :| | :        :|  :|:|   : ||
Db        782 KIAQVKPNNTGIRASVYEKTAKNGAKYADRTFYVTKERAHGNETYVLLNNTSHNIPLGWF 841

Qy         60 ----LNINN--EKSYVSKDITIKDQIQGGQQLDLSTLN-INVTGTHSNYYSGQSAITDFE 112
                ||: |  ::   ::  |:    |   :   | | : :|| :       |       :
Db        842 NVKDLNVQNLGKEVKTTQKYTVNKSNNGLSMVPWGTKNQVILTGNN----IAQGTFNATK 897

Qy        113 KAFPGSKI----TVDNTK---NTIDVTIP-------------------QGYGSY----N 141
              :   | :    |::|     |  |:| |                   | | |     |
Db        898 QVSVGKDVYLYGTINNRTGWVNAKDLTAPTAVKPTTSAAKDYNYTYVIKNGNGYYYVTPN 957

Qy        142 SFSINYKTKITNEQQ----KEFVNNSQAWYQEHGKEEVNGK--------------SFNHT 183
              | :  |  |  |||     || | | | ||  :||   |||                :| |
Db        958 SDTAKYSLKAFNEQPFAVVKEQVINGQTWY--YGKLS-NGKLAWIKSTDLAKELIKYNQT 1014

Qy        184 VHNINANAGIEGTVKGELKV--------------LKQDKDTK 211
               :|  | |:  :: : :|              :|   |||
Db       1015 GMTLNQVAQIQAGLQYKPQVQRVPGKWTDANFNDVKHAMDTK 1056


RESULT 14
C70203
DNA topoisomerase I (topA) homolog - Lyme disease spirochete
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 16-Jul-1999
C;Accession: C70203
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943; PMID:9403685
A;Accession: C70203
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-848 <KLE>
A;Cross-references: GB:AE001180; GB:AE000783; NID:g2688755; PIDN:AAC67161.1; PID:g268875
A;Experimental source: strain B31
C;Superfamily: bacterial type I DNA topoisomerase

  Query Match             8.6%;  Score 96.5;  DB 2;  Length 848;
  Best Local Similarity   22.8%;  Pred. No. 29;
  Matches   59;  Conservative   41;  Mismatches   84;  Indels   75;  Gaps   14;

Qy         19 ITSGNKSTNVTVHKSEAGTSSVFYY----------KTGDMLPEDTTHVRWFLNINNEKSY 68
              |  |:  : | :  ||  | : | |          | |    |   : :   ::   |: |
Db        457 IKKGDTFSIVKMKTSEHETKAPFRYTEASLVQKMEKEGIGRPSTYSTI---ISTLLEREY 513

Qy         69 VSK-DITIKDQIQGGQQLDL------STLNINVTGTH-----------------SNYYS 103
                | : |:   |:|   ::|        : :| |                   | :|:
Db        514 AFKLNNTLMPTIKGAAVINLLEKYFPVLIELNFTSNMEEKLDKIAIGKLDKIKYLSKFYN 573

Qy        104 GQSAITD-------------FEKAFPGSKITVDNTKNTIDVTIPQG-YGSY-----NSFS 144
              |:  : |             |       ||     ||:|: ||  | || |      :::|
Db        574 GKKGLKDTVMQLEPKIDSSEFRTVIESQKI---ENKNSINYTINIGKYGPYLIFKGHNYS 630

Qy        145 INYKTKITNEQQKE----FVNNSQAWYQEHGKEEVNGKS--FNHTVH-NI-----NANAG 192
              || || : |  :|:     :|  :       | : : | :  | :|:: ||     : :|
Db        631 INAKTPLENLYKKDEIEKIINEKELKPNILGVDPLTGLNVIFKNTIYGNIVQLGEDTHAP 690

Qy        193 IEGTVKG---ELKVLKQDK 208
               | | ||   :||::|  |
Db        691 QEYTKKGKPKKLKIIKAKK 709


RESULT 15
A53580
neurexin III beta precursor (clone pB794-7) - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 12-Apr-1995 #sequence_revision 12-Apr-1995 #text_change 09-Dec-2002
C;Accession: A53580
R;Ushkaryov, Y.A.; Hata, Y.; Ichtchenko, K.; Moomaw, C.; Afendis, S.; Slaughter, C.A.; S
J. Biol. Chem. 269, 11987-11992, 1994
A;Title: Conserved domain structure of beta-neurexins. Unusual cleaved signal sequences
A;Reference number: A53580; MUID:94216308; PMID:8163501
A;Accession: A53580
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-392 <USH>
A;Cross-references: GB:L27864
C;Superfamily: neurexin; EGF homology
C;Keywords: alternative splicing

  Query Match             8.6%;  Score 96;  DB 2;  Length 392;
  Best Local Similarity   19.6%;  Pred. No. 12;
  Matches   47;  Conservative   35;  Mismatches   74;  Indels   84;  Gaps   10;

Qy          5 HHHHHHGSDDKVATITSGNKSTNVTVHKS-------EAGTSSVFYYKTGDMLPEDTTHVR 57
              | || |||          : |  :::::|        || : :|    | :|
Db         52 HEHHFHGSK---------HHSVPISIYRSPVSLRGGHAGATYIFGKSGGLIL------YT 96

Qy         58 WFLN----INNEKSYVSKDITIKDQI-------------------QG--GQQLDLSTLNI 92
              |  |       :::  |    |:|| |                    ||  |   :: |::|
Db         97 WPANDRPSTRSDRLAVGFSTTVKDGILVRIDSAPGLGDFLQLHIEQGKIGVVFNIGTVDI 156

Qy         93 NVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIPQGYGSYNSFSIN--YKTK 150
              ::    :   |:  :  | :    : : |||                  : :|  | |
Db        157 SIKEERTPVNDGKYHVVRFTRNGGNATLQVDN-----------------WPVNEHYPTG 198

Qy        151 ITNEQQKEFVNNS---------QAWYQEHGKEEVNGKSFNHTVHNINANAGIEGTVKGEL 201
              |: :: : |             : | || |::          |: |  |   | |  || |
Db        199 NTDNERFQMVKQKIPFKYNRPVEEWLQEKGRQ--------LTIFNTQAQIAIGGKDKGRL 250


Search completed: July 27, 2004, 10:09:59
Job time : 17 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 27, 2004, 10:04:21 ; Search time 13 Seconds
(without alignments)
845.139 Million cell updates/sec

Title: US-09-813-820-4
Perfect score: 1122
Sequence: 1 MRGSHHHHHHGSDDKVATIT.....GIEGTVKGELKVLKQDKDTK 211

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1048	93.4	1183	1	CNA_STAAU	Q53654 staphylococ
2	106	9.4	1076	1	HSER_CAVPO	P70106 cavia porce
3	98	8.7	1256	1	ATL_STAAU	P52081 staphylococ
4	96.5	8.6	848	1	TOP1_BORBU	O51768 borrelia bu
5	96	8.6	456	1	NX3B_BOVIN	Q28143 bos taurus
6	96	8.6	704	1	MSN2_YEAST	P33748 saccharomyc
7	95	8.5	666	1	MUR2_ENTHR	P39046 enterococcu
8	95	8.5	2003	1	YDBA_ECOLI	P33666 escherichia
9	93.5	8.3	1296	1	BXG_CLOBO	Q60393 clostridium
10	92.5	8.2	711	1	TOP1_MYCPN	P78032 mycoplasma
11	91.5	8.2	238	1	OPAH_NEIGO	Q04884 neisseria g
12	91	8.1	463	1	GUN_BACSP	P29019 bacillus sp
13	91	8.1	578	1	LIPA_MYCPU	Q50274 mycoplasma
14	90.5	8.1	1169	1	C1GB_BACTZ	Q9zaz6 bacillus th
15	90	8.0	625	1	BSL1_TRIVA	Q8mti2 trichomonas
16	89.5	8.0	416	1	ELT1_CAEEL	P28515 caenorhabdi
17	89	7.9	599	1	LAC2_THACU	Q02075 thanatephor
18	89	7.9	858	1	CYAG_DICDI	Q03101 dictyosteli
19	88.5	7.9	520	1	AMY_BACME	P20845 bacillus me
20	88.5	7.9	576	1	DEAF_DROME	Q24180 drosophila
21	88.5	7.9	717	1	UBPB_YEAST	P36026 saccharomyc
22	87.5	7.8	382	1	Y077_METJA	Q60384 methanococc
23	87.5	7.8	523	1	RPN3_YEAST	P40016 saccharomyc
24	87.5	7.8	626	1	PC11_YEAST	P39081 saccharomyc
25	87.5	7.8	1435	1	DPO3_MYCPU	P47729 mycoplasma
26	87.5	7.8	2867	1	RBP2_PLAVB	Q00799 plasmodium
27	87	7.8	460	1	MURC_THETN	Q8r749 thermoanaer
28	87	7.8	668	1	PBPC_BACSU	P42971 bacillus su
29	87	7.8	762	1	SLAP_ACEKI	P22258 acetogenium
30	87	7.8	1310	1	YB35_SCHPO	O14340 schizosacch
31	87	7.8	1433	1	SUBF_BACSU	P16397 bacillus su
32	86.5	7.7	475	1	Y123_MYCPN	P75513 mycoplasma
33	86.5	7.7	678	1	YNC7_YEAST	P53968 saccharomyc
34	86.5	7.7	1694	1	IGA0_HAEIN	P44969 haemophilus
35	86.5	7.7	1702	1	IGA2_HAEIN	P45384 haemophilus
36	86	7.7	437	1	ACMA_LACLC	Q48603 lactococcus
37	86	7.7	617	1	Y237_BUCAI	P57331 buchnera ap
38	86	7.7	986	1	GUNZ_CLOSR	P23659 clostridium
39	85.5	7.6	664	1	AROB_FUSNN	Q8rf47 fusobacteri
40	85.5	7.6	850	1	PRSN_PIEBR	Q9gv36 pieris bras
41	85.5	7.6	938	1	PM15_CHLPN	Q9z883 chlamydia p
42	85.5	7.6	1146	1	AS10_YEAST	P48361 saccharomyc
43	85.5	7.6	1277	1	IF3X_YEAST	Q03690 saccharomyc
44	85	7.6	691	1	PPZ1_YEAST	P26570 saccharomyc
45	85	7.6	1088	1	VATA_CANTR	P38078 candida tro

ALIGNMENTS

RESULT 1
CNA_STAAU
ID CNA_STAAU STANDARD; PRT; 1183 AA.
AC Q53654;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen adhesin precursor.
GN CNA.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FDA 574;
RX MEDLINE=92165839; PubMed=1311320;
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
RA Lindberg M., Hoeoek M.;
RT "Molecular characterization and expression of a gene encoding a
RT Staphylococcus aureus collagen adhesin.";
RL J. Biol. Chem. 267:4766-4772(1992).
RN [2]
RP ERRATUM.
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
RA Lindberg M., Hoeoek M.;
RL J. Biol. Chem. 269:11672-11672(1994).
RN [3]
RP COLLAGEN-BINDING DOMAIN.
RC STRAIN=FDA 574;
RX MEDLINE=94032261; PubMed=8218209;
RA Patti J.M., Boles J.O., Hoeoek M.;
RT "Identification and biochemical characterization of the ligand
RT binding domain of the collagen adhesin from Staphylococcus aureus.";
RL Biochemistry 32:11428-11435(1993).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.
RX MEDLINE=97475225; PubMed=9334749;
RA Symersky J., Patti J.M., Carson M., House-Pompeo K., Teale M.,
RA Moore D., Jin L., Schneider A., DeLucas L.J., Hoeoek M.,
RA Narayana S.V.L.;
RT "Structure of the collagen-binding domain from a Staphylococcus
RT aureus adhesin.";
RL Nat. Struct. Biol. 4:833-838(1997).
CC -!- FUNCTION: MEDIATES ATTACHMENT OF STAPHYOCOCCAL CELLS TO
CC COLLAGEN-CONTAINING SUBSTRATA.
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (Potential).
CC ---

CC ---

```
DR   EMBL; M81736; AAA20874.1; -.
DR   PDB; 1AMX; 24-JUN-98.
DR   PDB; 1D2O; 27-SEP-00.
DR   PDB; 1D2P; 27-SEP-00.
DR   InterPro; IPR008966; Adhes_bact.
DR   InterPro; IPR008454; Cna_B.
DR   InterPro; IPR008970; Cna_B_unit.
DR   InterPro; IPR008456; Collagen_bind.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   Pfam; PF05738; Cna_B; 7.
DR   Pfam; PF05737; Collagen_bind; 1.
DR   TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR   PROSITE; PS50847; GRAM_POS_ANCHORING; FALSE_NEG.
KW   Cell wall; Peptidoglycan-anchor; Repeat; Signal; 3D-structure.
FT   SIGNAL       1     29       POTENTIAL.
FT   CHAIN       30   1154       COLLAGEN ADHESIN.
FT   PROPEP    1155   1183       REMOVED BY SORTASE (POTENTIAL).
FT   DOMAIN     151    318       COLLAGEN-BINDING.
FT   DOMAIN     533   1093       3 X 187 AA APPROXIMATE TANDEM REPEATS.
FT   DOMAIN    1093   1157       LYS/PRO-RICH (CELL WALL-SPANNING).
FT   REPEAT     533    719       B1.
FT   REPEAT     720    906       B2.
FT   REPEAT     907   1093       B3.
FT   SITE      1151   1155       LPXTG SORTING SIGNAL (POTENTIAL).
FT   MOD_RES   1154   1154       AMIDE-LINKED TO CELL WALL (POTENTIAL).
FT   STRAND     174    179
FT   TURN       182    183
FT   TURN       185    186
FT   STRAND     187    194
FT   TURN       196    197
FT   STRAND     201    201
FT   STRAND     205    211
FT   STRAND     215    228
FT   TURN       229    230
FT   STRAND     232    234
FT   HELIX      239    246
FT   TURN       248    249
FT   STRAND     251    255
FT   TURN       256    259
FT   STRAND     260    265
FT   HELIX      267    270
FT   TURN       271    272
FT   STRAND     273    283
FT   TURN       286    287
FT   STRAND     290    299
FT   STRAND     301    301
FT   TURN       302    303
FT   STRAND     307    311
FT   STRAND     314    317
SQ   SEQUENCE   1183 AA;  133066 MW;  B6A1CC072E575D76 CRC64;

  Query Match             93.4%;  Score 1048;  DB 1;  Length 1183;
  Best Local Similarity   100.0%;  Pred. No. 5.6e-70;
  Matches  200;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         12 SDDKVATITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSK 71
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        144 SDDKVATITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSK 203

Qy         72 DITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDV 131
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        204 DITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDV 263

Qy        132 TIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINANA 191
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        264 TIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINANA 323

Qy        192 GIEGTVKGELKVLKQDKDTK 211
              ||||||||||||||||||||
Db        324 GIEGTVKGELKVLKQDKDTK 343


RESULT 2
```

```
HSER_CAVPO
ID   HSER_CAVPO     STANDARD;      PRT;  1076 AA.
AC   P70106;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Heat-stable enterotoxin receptor precursor (GC-C) (Intestinal
DE   guanylate cyclase) (EC 4.6.1.2) (STA receptor) (Guanylyl cyclase C).
GN   GUCY2C OR GUC2C.
OS   Cavia porcellus (Guinea pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX   NCBI_TaxID=10141;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon;
RA   Kruhoeffer M., Cetin Y., Kaempf U., Forssmann W.-G.;
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: RECEPTOR FOR THE E.COLI HEAT-STABLE ENTEROTOXIN (E.COLI
CC       ENTEROTOXIN MARKEDLY STIMULATES THE ACCUMULATION OF CGMP IN
CC       MAMMALIAN CELLS EXPRESSING GC-C). ALSO ACTIVATED BY THE ENDOGENOUS
CC       PEPTIDE GUANYLIN.
CC   -!- CATALYTIC ACTIVITY: GTP = 3',5'-cyclic GMP + diphosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: Belongs to the adenylyl cyclase class-4/guanylyl
CC       cyclase family.
CC   -!- SIMILARITY: Contains 1 protein kinase-like domain.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z74734; CAA98989.1; -.
DR   HSSP; Q02846; 1AWL.
DR   InterPro; IPR001828; ANF_receptor.
DR   InterPro; IPR001054; G_cyclase.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF01094; ANF_receptor; 1.
DR   Pfam; PF00211; guanylate_cyc; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00044; CYCc; 1.
DR   PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
DR   PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Receptor; Transmembrane; Glycoprotein; Lyase; cGMP biosynthesis;
KW   Signal.
FT   SIGNAL       1     23       POTENTIAL.
FT   CHAIN       24   1076       HEAT-STABLE ENTEROTOXIN RECEPTOR.
FT   DOMAIN      24    433       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   434    457       POTENTIAL.
FT   DOMAIN     458   1076       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     492    752       PROTEIN KINASE-LIKE.
FT   DOMAIN     827    957       GUANYLATE CYCLASE.
FT   CARBOHYD    35     35       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    82     82       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   191    191       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   198    198       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   287    287       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   306    306       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   310    310       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   348    348       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   360    360       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   405    405       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1076 AA;  123119 MW;  9B53F16E05E80EB9 CRC64;

  Query Match              9.4%;  Score 106;  DB 1;  Length 1076;
```

```
  Best Local Similarity   19.6%;  Pred. No. 2.6;
  Matches   47;  Conservative   37;  Mismatches   78;  Indels   78;  Gaps   10;

Qy         24 KSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFLN-INNEKSYVSKDITIKDQIQGG 82
              |::|:         || ||  | |     :: |  |::| :    || |: :  |: ::|
Db        179 KASNLPFKSFSWNTSYVF--KNG----TESEHCFWYINALEAGVSYFSQVLGFKEMLRGN 232

Qy         83 QQLD---------------------LSTLNINVTGTH----------SNYYSGQSAITDF 111
              ::|                     : :| |: | |           :|||  ::|
Db        233 EELQKILKDPNRRSNVIVMCGTPQTMESLKIDWTATEDTVIILVDLFNNYYLEENATAP- 291

Qy        112 EKAFPGSKITVDNTKNTIDVTIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQAW----- 166
                  |  || : :|:| |  : |       | :: |  :|| : : |:
Db        292 -----------DYMKNVLVLTLPPGNSTIN-------TSLSKESLQEFSDFALAYLDGIL 333

Qy        167 --------YQEHGKEEVNGKSFNHTVHNINANAGI-------EGTVKGELKVLKQDKDTK 211
                    :  :| |       | |   |:           | :   : :|    |||
Db        334 LFGHMLKTFLRNG-ENTTAHKFAHAFRNLTFEGSTGPVTLDDSGDIDNTMVLLYTSVDTK 392


RESULT 3
ATL_STAAU
ID   ATL_STAAU      STANDARD;      PRT;  1256 AA.
AC   P52081;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Bifunctional autolysin precursor [Includes: N-acetylmuramoyl-L-alanine
DE   amidase (EC 3.5.1.28); Mannosyl-glycoprotein endo-beta-N-
DE   acetylglucosamidase (EC 3.2.1.96)].
GN   ATL.
OS   Staphylococcus aureus.
OC   Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX   NCBI_TaxID=1280;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 205-214 AND 776-792.
RC   STRAIN=RN450;
RX   MEDLINE=95116542; PubMed=7816834;
RA   Oshida T., Sugai M., Komatsuzawa H., Hong Y.-M., Suginaka H.,
RA   Tomasz A.;
RT   "A Staphylococcus aureus autolysin that has an N-acetylmuramoyl-L-
RT   alanine amidase domain and an endo-beta-N-acetylglucosaminidase
RT   domain: cloning, sequence analysis, and characterization.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:285-289(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 8325-4;
RA   Foster S.J.;
RL   Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ENDOHYDROLYSIS OF THE DI-N-ACETYLCHITOBIOSYL UNIT IN
CC       HIGH-MANNOSE GLYCOPEPTIDES AND GLYCOPROTEINS CONTAINING THE
CC       -[(MAN)5(GLCNAC)2]-ASN STRUCTURE. ONE N-ACETYL-D-GLUCOSAMINE
CC       RESIDUE REMAINS ATTACHED TO THE PROTEIN; THE REST OF THE
CC       OLIGOSACCHARIDE IS RELEASED INTACT.
CC   -!- CATALYTIC ACTIVITY: Hydrolyzes the link between N-acetylmuramoyl
CC       residues and L-amino acid residues in certain bacterial cell-wall
CC       glycopeptides.
CC   -!- CATALYTIC ACTIVITY: Endohydrolysis of the di-N-acetylchitobiosyl
CC       unit in high-mannose glycopeptides and glycoproteins containing
CC       the -[Man(GlcNAc)2]Asn-structure. One N-acetyl-D-glucosamine
CC       residue remains attached to the protein; the rest of the
CC       oligosaccharide is released intact.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- PTM: UNDERGOES PROTEOLYTIC PROCESSING TO GENERATE THE TWO
CC       EXTRACELLULAR LYTIC ENZYMES.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE N-
CC       ACETYLMURAMOYL-L-ALANINE AMIDASE FAMILY 2.
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 73 OF
CC       GLYCOSYL HYDROLASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D17366; BAA04185.1; -.
DR   EMBL; L41499; AAA99982.1; -.
DR   InterPro; IPR002502; Amidase_2.
DR   InterPro; IPR002901; Amidase_4.
DR   Pfam; PF01510; Amidase_2; 1.
DR   Pfam; PF01832; Amidase_4; 1.
DR   SMART; SM00644; Ami_2; 1.
DR   SMART; SM00047; LYZ2; 1.
KW   Cell wall; Hydrolase; Signal; Multifunctional enzyme; Repeat.
FT   SIGNAL       1     29       POTENTIAL.
FT   CHAIN       30   1256       BIFUNCTIONAL AUTOLYSIN.
FT   DOMAIN     199    775       N-ACETYLMURAMOYL-L-ALANINE AMIDASE.
FT   DOMAIN     776   1256       ENDO-BETA-N-ACETYLGLUCOSAMIDASE.
FT   REPEAT     425    589       1.
FT   REPEAT     596    758       2.
FT   REPEAT     770    932       3.
SQ   SEQUENCE   1256 AA;  137384 MW;  2BB76CAA292FDD20 CRC64;

  Query Match             8.7%;  Score 98;  DB 1;  Length 1256;
  Best Local Similarity   21.6%;  Pred. No. 12;
  Matches   61;  Conservative   39;  Mismatches   90;  Indels   92;  Gaps   16;

Qy         15 KVATITSGNKSTNVTVHKSEAGTSSVF----YYKTGD--------MLPEDTTH---VRWF 59
              |:| :   |     :|::  |   : :    :| | :         :|  :|:|   : ||
Db        790 KIAQVKPNNTGIRASVYEKTAKNGAKYADRTFYVTKERAHGNETYVLLNNTSHNIPLGWF 849

Qy         60 ----LNINN--EKSYVSKDITIKDQIQGGQQLDLSTLN-INVTGTHSNYYSGQSAITDFE 112
                  ||: |  ::   ::  |:     |   :   | | : :|| :      |      :
Db        850 NVKDLNVQNLGKEVKTTQKYTVNKSNNGLSMVPWGTKNQVILTGNN----IAQGTFNATK 905

Qy        113 KAFPGSKI----TVDNTK---NTIDVTIP--------------------QGYGSY----N 141
              :   |  :    |::|     |  |:| |                        | | |    |
Db        906 QVSVGKDVYLYGTINNRTGWVNAKDLTAPTAVKPTTSAAKDYNYTYVIKNGNGYYYVTPN 965

Qy        142 SFSINYKTKITNEQQ----KEFVNNSQAWYQEHGKEEVNGK--------------SFNHT 183
              | :  |  |  |||     || | | | ||  :||   |||              :| |
Db        966 SDTAKYSLKAFNEQPFAVVKEQVINGQTWY--YGKLS-NGKLAWIKSTDLAKELIKYNQT 1022

Qy        184 VHNINANAGIEGTVKGELKV--------------LKQDKDTK 211
                 :|  | |:  :: : :|               :|   |||
Db       1023 GMTLNQVAQIQAGLQYKPQVQRVPGKWTDAKFNDVKHAMDTK 1064


RESULT 4
TOP1_BORBU
ID   TOP1_BORBU     STANDARD;      PRT;   848 AA.
AC   O51768;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   DNA topoisomerase I (EC 5.99.1.2) (Omega-protein) (Relaxing enzyme)
DE   (Untwisting enzyme) (Swivelase).
GN   TOPA OR BB0828.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35210 / B31;
RX   MEDLINE=98065943; PubMed=9403685;
RA   Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA   Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA   Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA   Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA   van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA   Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA   Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA   Smith H.O., Venter J.C.;
```

```
RT   "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT   burgdorferi.";
RL   Nature 390:580-586(1997).
CC   -!- FUNCTION: The reaction catalyzed by topoisomerases leads to the
CC       conversion of one topological isomer of DNA to another.
CC   -!- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded
CC       DNA, followed by passage and rejoining.
CC   -!- SUBUNIT: Monomer (By similarity).
CC   -!- MISCELLANEOUS: When a topoisomerase transiently breaks a DNA
CC       backbone bond, it simultaneously forms a protein-DNA link, in
CC       which a tyrosyl oxygen in the enzyme is joined to a DNA phosphorus
CC       at one end of the enzyme-severed DNA strand.
CC   -!- SIMILARITY: Belongs to the prokaryotic type I/III topoisomerase
CC       family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001180; AAC67161.1; -.
DR   PIR; C70203; C70203.
DR   HSSP; P06612; 1ECL.
DR   TIGR; BB0828; -.
DR   InterPro; IPR005733; DNA_topI_bact.
DR   InterPro; IPR000380; DNA_tpisomrase.
DR   InterPro; IPR003601; DNAtopI_ATP_bind.
DR   InterPro; IPR003602; DNAtopI_DNA_bind.
DR   InterPro; IPR006171; Toprim_dom.
DR   InterPro; IPR006154; Toprim_sub.
DR   Pfam; PF01131; Topoisom_bac; 1.
DR   Pfam; PF01751; Toprim; 1.
DR   PRINTS; PR00417; PRTPISMRASEI.
DR   SMART; SM00437; TOP1Ac; 1.
DR   SMART; SM00436; TOP1Bc; 1.
DR   SMART; SM00493; TOPRIM; 1.
DR   TIGRFAMs; TIGR01051; topA_bact; 1.
DR   PROSITE; PS00396; TOPOISOMERASE_I_PROK; 1.
KW   Isomerase; Topoisomerase; DNA-binding; Complete proteome.
FT   ACT_SITE    324    324       DNA CLEAVAGE (BY SIMILARITY).
SQ   SEQUENCE   848 AA;  97663 MW;  37DEA5FB1E12FA2C CRC64;

  Query Match             8.6%;  Score 96.5;  DB 1;  Length 848;
  Best Local Similarity   22.8%;  Pred. No. 9.9;
  Matches   59;  Conservative   41;  Mismatches   84;  Indels   75;  Gaps   14;

Qy         19 ITSGNKSTNVTVHKSEAGTSSVFYY----------KTGDMLPEDTTHVRWFLNINNEKSY 68
              |  |:  : | :  ||  | : | |          | |   |   : :   ::   |: |
Db        457 IKKGDTFSIVKMKTSEHETKAPFRYTEASLVQKMEKEGIGRPSTYSTI---ISTLLEREY 513

Qy         69 VSK-DITIKDQIQGGQQLDL------STLNINVTGTH-----------------SNYYS 103
                | : |:   |:|   ::|        : :| |                     | :|:
Db        514 AFKLNNTLMPTIKGAAVINLLEKYFPVLIELNFTSNMEEKLDKIAIGKLDKIKYLSKFYN 573

Qy        104 GQSAITD-------------FEKAFPGSKITVDNTKNTIDVTIPQG-YGSY-----NSFS 144
              |:  : |             |       ||     ||:|: ||  | || |     :::|
Db        574 GKKGLKDTVMQLEPKIDSSEFRTVIESQKI---ENKNSINYTINIGKYGPYLIFKGHNYS 630

Qy        145 INYKTKITNEQQKE----FVNNSQAWYQEHGKEEVNGKS--FNHTVH-NI-----NANAG 192
              || || : |  :|:     :|  :      | : : | :  | :|:: ||      : :|
Db        631 INAKTPLENLYKKDEIEKIINEKELKPNILGVDPLTGLNVIFKNTIYGNIVQLGEDTHAP 690

Qy        193 IEGTVKG---ELKVLKQDK 208
                | | ||   :||::|  |
Db        691 QEYTKKGKPKKLKIIKAKK 709


RESULT 5
NX3B_BOVIN
ID   NX3B_BOVIN     STANDARD;      PRT;   456 AA.
AC   Q28143; Q28144;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Neurexin 3-beta precursor (Neurexin III-beta).
GN   NRXN3.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3 AND 4), AND SEQUENCE OF
RP   N-TERMINUS.
RC   TISSUE=Brain;
RX   MEDLINE=94216308; PubMed=8163501;
RA   Ushkaryov Y.A., Hata Y., Ichtchenko K., Moomaw C., Afendis S.,
RA   Slaughter C.A., Suedhof T.C.;
RT   "Conserved domain structure of beta-neurexins. Unusual cleaved signal
RT   sequences in receptor-like neuronal cell-surface proteins.";
RL   J. Biol. Chem. 269:11987-11992(1994).
CC   -!- FUNCTION: Neuronal cell surface protein that may be involved in
CC       cell recognition and cell adhesion.
CC   -!- SUBUNIT: The cytoplasmic C-terminal region binds to CASK (By
CC       similarity). Binds to neuroligins NLGN1, NLGN2 and NLGN3 (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC       Isoforms 3 and 4 seems to be secreted.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=4;
CC         Comment=Additional isoforms seem to exist. There is a
CC         combination of two alternatively spliced domains at sites 4 and
CC         5, which seem to be used independently. Experimental
CC         confirmation may be lacking for some isoforms;
CC       Name=1;
CC         IsoId=Q28143-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q28143-2; Sequence=VSP_003533;
CC       Name=3;
CC         IsoId=Q28143-3; Sequence=VSP_003534;
CC         Note=Lacks the transmembrane domain;
CC       Name=4;
CC         IsoId=Q28143-4; Sequence=VSP_003533, VSP_003534;
CC         Note=Lacks the transmembrane domain;
CC   -!- SIMILARITY: Contains 1 laminin G-like domain.
CC   -!- SIMILARITY: Belongs to the neurexin family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L27869; AAA19907.1; -.
DR   EMBL; L27869; AAA19908.1; ALT_INIT.
DR   PIR; A53580; A53580.
DR   PIR; B53580; B53580.
DR   HSSP; Q63373; 1C4R.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003585; Neurexin-like.
DR   Pfam; PF00054; laminin_G; 1.
DR   SMART; SM00294; 4.1m; 1.
DR   SMART; SM00282; LamG; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 1.
KW   Signal; Transmembrane; Cell adhesion; Glycoprotein;
KW   Alternative splicing.
FT   SIGNAL        1     35
FT   CHAIN        36    456       NEUREXIN 3-BETA.
FT   DOMAIN       36    381       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    382    402       POTENTIAL.
```

```
FT   DOMAIN       403    456       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        82    282       LAMININ G-LIKE.
FT   DOMAIN        41     45       POLY-SER.
FT   DOMAIN       311    314       POLY-THR.
FT   DOMAIN       317    320       POLY-THR.
FT   DOMAIN       389    392       POLY-ALA.
FT   CARBOHYD     181    181       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     279    279       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     323    323       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC     198    227       Missing (in isoform 2 and isoform 4).
FT                                 /FTId=VSP_003533.
FT   VARSPLIC     357    456       ANPTEPGVRRVPGASEVIRESSSTTGMVVGIVAAAALCILI
FT                                 LLYAMYKYRNRDEGSYQVDETRNYISNSAQSNGTLLKEKPP
FT                                 SSKGGHKKQKNKDKEYYV -> ARSSNAARSLRAALTWTWR
FT                                 LTYTFTPIIFISCVVHS (in isoform 3 and
FT                                 isoform 4).
FT                                 /FTId=VSP_003534.
SQ   SEQUENCE   456 AA;  50244 MW;  E1B5467089A8CF2C CRC64;

  Query Match             8.6%;  Score 96;  DB 1;  Length 456;
  Best Local Similarity   19.6%;  Pred. No. 5;
  Matches   47;  Conservative   35;  Mismatches   74;  Indels   84;  Gaps   10;

Qy          5 HHHHHHGSDDKVATITSGNKSTNVTVHKS-------EAGTSSVFYYKTGDMLPEDTTHVR 57
              | || |||          : |  :::::|        || : :|    | :|
Db         52 HEHHFHGSK---------HHSVPISIYRSPVSLRGGHAGATYIFGKSGGLIL------YT 96

Qy         58 WFLN----INNEKSYVSKDITIKDQI------------------QG--GQQLDLSTLNI 92
              | |      :::  |    |:|| |                  ||  |   :: |::|
Db         97 WPANDRPSTRSDRLAVGFSTTVKDGILVRIDSAPGLGDFLQLHIEQGKIGVVFNIGTVDI 156

Qy         93 NVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIPQGYGSYNSFSIN--YKTK 150
              ::    :    |:  :  | :    : : |||                 : :|  | |
Db        157 SIKEERTPVNDGKYHVVRFTRNGGNATLQVDN-----------------WPVNEHYPTG 198

Qy        151 ITNEQQKEFVNNS---------QAWYQEHGKEEVNGKSFNHTVHNINANAGIEGTVKGEL 201
               |: :: : |           : | || |::         |: |  |   | |  || |
Db        199 NTDNERFQMVKQKIPFKYNRPVEEWLQEKGRQ--------LTIFNTQAQIAIGGKDKGRL 250


RESULT 6
MSN2_YEAST
ID   MSN2_YEAST     STANDARD;      PRT;   704 AA.
AC   P33748;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein MSN2 (Multicopy suppressor of SNF1 protein 2).
GN   MSN2 OR YMR037C OR YM9532.02C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=93309420; PubMed=8321194;
RA   Estruch F., Carlson M.;
RT   "Two homologous zinc finger genes identified by multicopy suppression
RT   in a SNF1 protein kinase mutant of Saccharomyces cerevisiae.";
RL   Mol. Cell. Biol. 13:3872-3881(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=97313268; PubMed=9169872;
RA   Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA   Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA   Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA   Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   XIII.";
RL   Nature 387:90-93(1997).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=96208509; PubMed=8641288;
RA   Martinez-Pastor M.T., Marchler G., Schueller C., Marchler-Bauer A.,
RA   Ruis H., Estruch F.;
RT   "The Saccharomyces cerevisiae zinc finger proteins Msn2p and Msn4p
RT   are required for transcriptional induction through the stress
RT   response element (STRE).";
RL   EMBO J. 15:2227-2235(1996).
RN   [4]
RP   NUCLEOCYTOPLASMIC SHUTTLING.
RX   MEDLINE=22628141; PubMed=12732613;
RA   Jacquet M., Renault G., Lallet S., De Mey J., Goldbeter A.;
RT   "Oscillatory nucleocytoplasmic shuttling of the general stress
RT   response transcriptional activators Msn2 and Msn4 in Saccharomyces
RT   cerevisiae.";
RL   J. Cell Biol. 161:497-505(2003).
CC   -!- FUNCTION: POSITIVE TRANSCRIPTIONAL FACTOR THAT ACTS AS A COMPONENT
CC       OF THE STRESS RESPONSIVE SYSTEM. RECOGNIZES AND BINDS TO THE
CC       STRESS RESPONSE ELEMENT (STRE) WHICH IS INVOLVED IN THE RESPONSE
CC       TO VARIOUS FORMS OF STRESS (HEAT, OXIDATIVE, OSMOTIC, ETC.).
CC       INVOLVED IN THE REGULATION OF THE CTT1, DDR2, HSP12 GENES.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear.
CC   -!- SIMILARITY: Contains 2 C2H2-type zinc fingers.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L08838; AAA34806.1; -.
DR   EMBL; Z48502; CAA88403.1; -.
DR   PIR; S39004; S39004.
DR   HSSP; P08047; 1SP1.
DR   GermOnline; 142706; -.
DR   TRANSFAC; T01257; -.
DR   SGD; S0004640; MSN2.
DR   GO; GO:0005829; C:cytosol; IDA.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0006950; P:response to stress; IMP.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 2.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 2.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 2.
KW   DNA-binding; Nuclear protein; Zinc-finger; Metal-binding; Activator;
KW   Transcription regulation; Repeat.
FT   DOMAIN         1    109       ASP-RICH (ACIDIC).
FT   DOMAIN       260    279       ASP-RICH (ACIDIC).
FT   ZN_FING      647    665       C2H2-TYPE 1.
FT   ZN_FING      676    698       C2H2-TYPE 2.
SQ   SEQUENCE   704 AA;  77860 MW;  EDF6F07446819DF1 CRC64;

  Query Match             8.6%;  Score 96;  DB 1;  Length 704;
  Best Local Similarity   25.4%;  Pred. No. 8.5;
  Matches   64;  Conservative   31;  Mismatches   97;  Indels   60;  Gaps   13;

Qy         12 SDDKVATITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSK 71
              ||    : :  | ::| | : :    || |     | : :|       || ::: : :|:
Db        229 SDTNSYSNSISNSNSNSTGNLN----SSYFNSLNIDSMLDDYVSSDLLLNDDDDDTNLSR 284

Qy         72 ----DITIKDQIQG--------GQQLDL----------STLNINVTGTHSNYYSGQSAIT 109
                  |: | :|             |||            |:|:| |:|   |  |  |
Db        285 RRFSDV-ITNQFPSMTNSRNSISHSLDLWNHPKINPSNRNTNLNIT-TNSTSSSNASPNT 342

Qy        110 DFEKAFPGSKITVDNTKN---TIDVTIPQGYGSYN--------------------SFSI 145
                 |   | |   | ||   |||  : |   ||                      ||
Db        343 TTMNANADSNI-AGNPKNNDATIDNELTQILNEYNMNFNDNLGTSTSGKNKSACPSSFDA 401

Qy        146 NYKTKITNEQQ-KEFVNNSQ--AWYQEHGKEEVNGKSFNHTVHNINANAG---IEGTVKG 199
```

```
              |  |||   || :: :|  |       |     | ||||  :::    |    |: ::
Db        402 NAMTKINPSQQLQQQLNRVQHKQLTSSHNNSSTNMKSFNSDLYSRRQRASLPIIDDSLSY 461

Qy        200 ELKVLKQDKDTK 211
              :| | |||:| |
Db        462 DL-VNKQDEDPK 472


RESULT 7
MUR2_ENTHR
ID   MUR2_ENTHR     STANDARD;      PRT;   666 AA.
AC   P39046;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Muramidase-2 precursor (EC 3.2.1.17) (1,4-beta-N-
DE   acetylmuramoylhydrolase) (Peptidoglycan hydrolase) (Pg-hydrolase-2)
DE   (Lysosyme).
OS   Enterococcus hirae.
OC   Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX   NCBI_TaxID=1354;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 50-73.
RC   STRAIN=ATCC 9790;
RX   MEDLINE=92165737; PubMed=1347040;
RA   Chu C.-P., Kariyama R., Daneo-Moore L., Shockman G.D.;
RT   "Cloning and sequence analysis of the muramidase-2 gene from
RT   Enterococcus hirae.";
RL   J. Bacteriol. 174:1619-1625(1992).
RN   [2]
RP   FUNCTION.
RC   STRAIN=ATCC 9790;
RX   MEDLINE=89327152; PubMed=2753858;
RA   Dolinger D.L., Daneo-Moore L., Shockman G.D.;
RT   "The second peptidoglycan hydrolase of Streptococcus faecium ATCC 9790
RT   covalently binds penicillin.";
RL   J. Bacteriol. 171:4355-4361(1989).
RN   [3]
RP   FUNCTION.
RC   STRAIN=ATCC 9790;
RA   Del Mar Lleo M., Canepari P., Satta G.;
RT   "Thermosensitive cell growth mutants of Enterococcus hirae that
RT   elongate at non-permissive temperature are stimulated to divide by
RT   parental autolytic enzymes.";
RL   J. Gen. Microbiol. 139:3099-3117(1993).
CC   -!- FUNCTION: MAY WORK IN CONCERT WITH AND POTENTIATE THE PROCESSIVE
CC       HYDROLYTIC ACTION OF MURAMIDASE-1, WHICH REQUIRES BINDING OF THE
CC       ENZYME TO NONREDUCING ENDS OF GLYCAN CHAINS. HYDROLYSIS IN THE
CC       MIDST OF GLYCAN CHAINS WOULD INCREASE THE NUMBER OF BINDING SITES
CC       FOR MURAMIDASE-1. MAY FUNCTION IN FACILITATING SEPTUM FORMATION
CC       AND CELL SEPARATION. ACTIVE ON M.LUTEUS CELL WALLS AND ON E.HIRAE
CC       CELL WALL FRACTIONS, BUT NOT ACTIVE ON E.HIRAE INTACT CELL WALLS.
CC       CAN COVALENTLY BIND PENICILLIN.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages between N-
CC       acetyl-D-glucosamine and N-acetylmuramic acid in peptidoglycan
CC       heteropolymers of the prokaryotes cell walls.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- DOMAIN: LysM repeats are thought to be involved in peptidoglycan
CC       binding.
CC   -!- SIMILARITY: Belongs to family 73 of glycosyl hydrolases.
CC   -!- SIMILARITY: Contains 6 LysM repeats.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M77639; AAA24776.1; -.
DR   PIR; A42296; A42296.
DR   HSSP; P22629; 1SWD.
DR   InterPro; IPR002901; Amidase_4.
DR   InterPro; IPR000423; Flag_FlgJ.
DR   InterPro; IPR002482; LysM.
DR   InterPro; IPR000437; Prok_lipoprot_S.
DR   Pfam; PF01832; Amidase_4; 1.
DR   Pfam; PF01476; LysM; 6.
DR   PRINTS; PR01002; FLGFLGJ.
DR   SMART; SM00257; LysM; 6.
DR   SMART; SM00047; LYZ2; 1.
KW   Hydrolase; Glycosidase; Bacteriolytic enzyme; Cell wall;
KW   Cell division; Septation; Repeat; Signal.
FT   SIGNAL       1     49
FT   CHAIN       50    666       MURAMIDASE-2.
FT   REPEAT     257    299       LYSM 1.
FT   REPEAT     338    380       LYSM 2.
FT   REPEAT     414    456       LYSM 3.
FT   REPEAT     489    531       LYSM 4.
FT   REPEAT     565    607       LYSM 5.
FT   REPEAT     623    665       LYSM 6.
SQ   SEQUENCE   666 AA;  70670 MW;  FF0A7FAFCD810BA3 CRC64;

  Query Match             8.5%;  Score 95;  DB 1;  Length 666;
  Best Local Similarity   24.0%;  Pred. No. 9.4;
  Matches   49;  Conservative   28;  Mismatches   81;  Indels   46;  Gaps   11;

Qy         20 TSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTH---------VRWFLNINNEKSYVS 70
              ||||   : |       | : :  |:|| :    :|          : |  || |   |
Db        242 TSGNSGGSATT------TGTTYTVKSGDSV-WGISHSFGITMAQLIEW-NNIKNNFIYPG 293

Qy         71 KDITIKDQIQGGQQLDLSTLNI-------NVTG-THSNYYSGQSAITDFEKAFPGSKITV 122
              : :|||    |||    || |         | :| |:::  :||:           |:| ||
Db        294 QKLTIK----GGQSAGSSTTNTGNNASSGNTSGNTNTSGSTGQAT---------GAKYTV 340

Qy        123 DNTKNTIDVTIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSFNH 182
               :  :   :    |       |:|   :  |  :  ||   |       |    :| : |
Db        341 KSGDSVWKIANDHG------ISMNQLIE-WNNIKNNFVYPGQQLVVSKGSSSASGSTSNT 393

Qy        183 TVHNINANAGIEG-TVKGELKVLK 205
              :  | ::|    | |  |    :|
Db        394 STGNTSSNTANTGSTTSGSTYTVK 417


RESULT 8
YDBA_ECOLI
ID   YDBA_ECOLI     STANDARD;      PRT;  2003 AA.
AC   P33666; P76087; P76088; P76856; P76857; P76859;
DT   01-FEB-1994 (Rel. 28, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein ydbA.
GN   YDBA OR B1401/B1405.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97251357; PubMed=9097039;
RA   Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA   Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA   Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA   Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
```

```
RA   Sampei G., Seki Y., Sivasundaram S., Tagami H., Takeda J.,
RA   Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT   "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 28.0-40.1 min region on the linkage map.";
RL   DNA Res. 3:363-377(1996).
RN   [3]
RP   SEQUENCE OF 464-2003 FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=92190338; PubMed=1665988;
RA   Moszer I., Glaser P., Danchin A.;
RT   "Multiple IS insertion sequences near the replication terminus in
RT   Escherichia coli K-12.";
RL   Biochimie 73:1361-1374(1991).
CC   -!- SIMILARITY: TO S.TYPHIMURIUM ORF NEAR CYSG (AC P25928).
CC   -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE GENE CODING FOR
CC       THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2D/IS30C ELEMENT
CC       BETWEEN AMINO ACIDS 839 AND 840.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000237; AAC74483.1; ALT_SEQ.
DR   EMBL; AE000237; AAC74487.1; ALT_SEQ.
DR   EMBL; D90778; BAA15009.1; ALT_SEQ.
DR   EMBL; D90778; BAA18880.1; ALT_SEQ.
DR   EMBL; D90779; BAA18881.1; ALT_SEQ.
DR   EMBL; X62680; -; NOT_ANNOTATED_CDS.
DR   EcoGene; EG11307; ydbA.
KW   Hypothetical protein; Complete proteome.
FT   CONFLICT    489    489       I -> V (IN REF. 2).
FT   CONFLICT    495    495       I -> V (IN REF. 2).
SQ   SEQUENCE   2003 AA;  205949 MW;  B83A12C8B53220EE CRC64;

  Query Match             8.5%;  Score 95;  DB 1;  Length 2003;
  Best Local Similarity   25.1%;  Pred. No. 37;
  Matches   69;  Conservative   30;  Mismatches   92;  Indels   84;  Gaps   16;

Qy          5 HHHHHHG----------SDDKVATITSGNK------------------STNVTVHKSEA 35
              ||| ::           |||     | |                  : :| : |:|
Db         94 HHHRNNSPLPPTPPDDESDDTPVPPTPGGDEIIPDDPPDDTPTPPKPVSFNNDVILDKTEK 153

Qy         36 GTS---SVFYY---KTGDMLPEDTTHVRWFLNI------NNEKSY--VSKDITIKDQIQG 81
                :   ||| |     | :  :|:   :  :|:      ||  :   || |   | |
Db        154 TLTIRDSVFTYTENADGTISLQDSNGRKATINLWQIDEANNTVALEGVSADGATKWQYNH 213

Qy         82 GQQL----DLSTLNIN---------VTGTHSNYYSGQSAITDFEK---------AFPGSK 119
                :|     | :|:| |           |||  |  :|:  | | :             |
Db        214 NGELVITGDNATVNNNGKTTVDGKDSTGTEINGNNGK-VIQDGDLDVSGGGHGIDITGDS 272

Qy        120 ITVDNTKNTIDVTIPQGYG---SYNSFSINY--KTKITNEQQKEFVNNSQAWYQEHGKEE 174
              |||| | |: || |:  |     :   :|   :: |||      :|   |    :||
Db        273 ATVDN-KGTMTVTDPESMGIQIDGDKAIVNNEGESTITNGGTGTQINGDDATANNNGKTT 331

Qy        175 VNGKSFNHTVHNINANAGIEGTVKGELKVLKQDKD 209
              |:||    |   || | |          ||: || |
Db        332 VDGKDSTGT--EINGNNG---------KVI-QDGD 354


RESULT 9
BXG_CLOBO
ID   BXG_CLOBO      STANDARD;      PRT;  1296 AA.
AC   Q60393;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type G precursor (EC 3.4.24.69) (BoNT/G)
DE   (Bontoxilysin G).
GN   BOTG.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=113 / 30;
RX   MEDLINE=94092745; PubMed=8268233;
RA   Campbell K., Collins M.D., East A.K.;
RT   "Nucleotide sequence of the gene coding for Clostridium botulinum
RT   (Clostridium argentinense) type G neurotoxin: genealogical comparison
RT   with other clostridial neurotoxins.";
RL   Biochim. Biophys. Acta 1216:487-491(1993).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X74162; CAA52275.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    441       BOTULINUM NEUROTOXIN G, LIGHT-CHAIN.
FT   CHAIN       442   1296       BOTULINUM NEUROTOXIN G, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    435    449       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1296 AA;  149013 MW;  DC8E47E15F665C31 CRC64;

  Query Match             8.3%;  Score 93.5;  DB 1;  Length 1296;
  Best Local Similarity   23.0%;  Pred. No. 28;
  Matches   50;  Conservative   32;  Mismatches   66;  Indels   69;  Gaps   12;

Qy         47 DMLP-------EDTTHVRWFLNINNEKSYVSKDITIKDQIQGGQQLDLS----TLNINVT 95
                | :|       :||  :: |   ||  | :| :  :    :||: :| |     |:|:
Db        838 DSIPFDLSLYTKDTILIQVF---NNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSD 894

Qy         96 GTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIPQGYGS-YNSFSINY------- 147
                ::  :||  : : |      | ||   :|  :       | | :::||||:
Db        895 VIFNDIGNGQFKLNNSE----NSNITAHQSKFVV-------YDSMFDNFSINFWVRTPKY 943
```

```
Qy        148 -KTKITNEQQKEF-----VNNSQAW-YQEHGKE------EVNGKS--------------- 179
                  |    | |:     : |   |      |         :|| ||
Db        944 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 1003

Qy        180 -----FNHTVHNI---NANAGIEGTVKGELKVLKQDK 208
                  |: |: |    |||  | |::|   |:| |:
Db       1004 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDR 1040


RESULT 10
TOP1_MYCPN
ID   TOP1_MYCPN     STANDARD;      PRT;   711 AA.
AC   P78032;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   DNA topoisomerase I (EC 5.99.1.2) (Omega-protein) (Relaxing enzyme)
DE   (Untwisting enzyme) (Swivelase).
GN   TOPA OR MPN261 OR MP572.
OS   Mycoplasma pneumoniae.
OC   Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2104;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29342 / M129;
RX   MEDLINE=97105885; PubMed=8948633;
RA   Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA   Herrmann R.;
RT   "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT   pneumoniae.";
RL   Nucleic Acids Res. 24:4420-4449(1996).
CC   -!- FUNCTION: The reaction catalyzed by topoisomerases leads to the
CC       conversion of one topological isomer of DNA to another.
CC   -!- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded
CC       DNA, followed by passage and rejoining.
CC   -!- SUBUNIT: Monomer (By similarity).
CC   -!- MISCELLANEOUS: When a topoisomerase transiently breaks a DNA
CC       backbone bond, it simultaneously forms a protein-DNA link, in
CC       which a tyrosyl oxygen in the enzyme is joined to a DNA phosphorus
CC       at one end of the enzyme-severed DNA strand.
CC   -!- SIMILARITY: Belongs to the prokaryotic type I/III topoisomerase
CC       family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000056; AAB96220.1; -.
DR   PIR; S73898; S73898.
DR   HSSP; P06612; 1ECL.
DR   InterPro; IPR005733; DNA_topI_bact.
DR   InterPro; IPR000380; DNA_tpisomrase.
DR   InterPro; IPR003601; DNAtopI_ATP_bind.
DR   InterPro; IPR003602; DNAtopI_DNA_bind.
DR   InterPro; IPR006171; Toprim_dom.
DR   InterPro; IPR006154; Toprim_sub.
DR   Pfam; PF01131; Topoisom_bac; 1.
DR   Pfam; PF01751; Toprim; 1.
DR   Pfam; PF01396; zf-C4_Topoisom; 2.
DR   PRINTS; PR00417; PRTPISMRASEI.
DR   SMART; SM00437; TOP1Ac; 1.
DR   SMART; SM00436; TOP1Bc; 1.
DR   SMART; SM00493; TOPRIM; 1.
DR   TIGRFAMs; TIGR01051; topA_bact; 1.
DR   PROSITE; PS00396; TOPOISOMERASE_I_PROK; 1.
KW   Isomerase; Topoisomerase; DNA-binding; Zinc-finger; Metal-binding;
KW   Repeat; Complete proteome.
FT   ZN_FING     624    652       C4-TYPE 1.
FT   ZN_FING     673    702       C4-TYPE 2.
FT   ACT_SITE    340    340       DNA CLEAVAGE (BY SIMILARITY).
SQ   SEQUENCE   711 AA;  81965 MW;  16B863588B06582E CRC64;

  Query Match             8.2%;  Score 92.5;  DB 1;  Length 711;
  Best Local Similarity   25.4%;  Pred. No. 16;
  Matches   49;  Conservative   32;  Mismatches   65;  Indels   47;  Gaps   12;

Qy         54 THVRWFLNINNE-----KSYV---------------SKDITIK-DQIQGGQQLDLSTLNI 92
              ||:| |:|  |:     || |                | |: |  |:|: | :     : |
Db        436 THIR-FINDGNKFYASSKSLVFDGYRKIYEHFENKESNDLYIDLDKIRVGDRFMAKDIKI 494

Qy         93 NVTGTH-SNYYSGQSAITDFEKAFPGSKITVDNTKNTI-DVTIPQGYGSYN--SFSINYK 148
                 || :  |:  | |   ||    | |   :| ||:  | : :|| | |  :| :
Db        495 TARQTHPAARYTQASLIEALEK----SNIGRPSTYNTMASVNLDRGYASLNKHAFHVTQL 550

Qy        149 TKITNEQ---------QKEFVNNSQAWYQEHGKEEV--NGKSFNHTVHNINANAGIE-GT 196
               :  ||:          |||  |       |   :|:  | |::   : :  :|   |
Db        551 GEQVNEELSKHFGKIINKEFTKN-----MEKSLDEIAENKKNYQEFLRDFWSNFKEEVKL 605

Qy        197 VKGELKVLKQDKD 209
               :| :: :|::|:
Db        606 AEGSIQRVKKEKE 618


RESULT 11
OPAH_NEIGO
ID   OPAH_NEIGO     STANDARD;      PRT;   238 AA.
AC   Q04884;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Opacity protein OPA60 precursor (Fragment).
GN   OPAH.
OS   Neisseria gonorrhoeae.
OC   Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC   Neisseriaceae; Neisseria.
OX   NCBI_TaxID=485;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MS11 / F3;
RX   MEDLINE=93178439; PubMed=8440254;
RA   Kupsch E.-M., Knepper B., Kuroki T., Heuer I., Meyer T.F.;
RT   "Variable opacity (Opa) outer membrane proteins account for the cell
RT   tropisms displayed by Neisseria gonorrhoeae for human leukocytes and
RT   epithelial cells.";
RL   EMBO J. 12:641-650(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MS11 / V18;
RX   MEDLINE=92114767; PubMed=1815562;
RA   Bhat K.S., Gibbs C.P., Barrera O., Morrison S.G., Jaehnig F.,
RA   Stern A., Kupsch E.-M., Meyer T.F., Swanson J.;
RT   "The opacity proteins of Neisseria gonorrhoeae strain MS11 are
RT   encoded by a family of 11 complete genes.";
RL   Mol. Microbiol. 5:1889-1901(1991).
RN   [3]
RP   ERRATUM.
RX   MEDLINE=92261323; PubMed=1584024;
RA   Bhat K.S., Gibbs C.P., Barrera O., Morrison S.G., Jaehnig F.,
RA   Stern S., Kupsch E.-M., Meyer T.F., Swanson J.;
RL   Mol. Microbiol. 6:1073-1076(1992).
CC   -!- FUNCTION: IMPLICATED IN A NUMBER OF ADHERENCE FUNCTIONS. OPA
CC       PROTEINS ARE IMPLICATED IN PATHOGENESIS AND ARE SUBJECT TO PHASE
CC       VARIATION.
CC   -!- SUBCELLULAR LOCATION: Outer membrane.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ----------------------------------------------------------------------
DR   EMBL; Z18939; CAA79372.1; -.
DR   EMBL; X60711; CAA43121.1; -.
DR   PIR; S16619; S16619.
DR   InterPro; IPR003394; Porin_opacity.
DR   Pfam; PF02462; Opacity; 1.
KW   Outer membrane; Multigene family; Signal.
FT   NON_TER       1      1
FT   SIGNAL       <1      1       POTENTIAL.
FT   CHAIN         2   >238       OPACITY PROTEIN OPA60.
FT   VARIANT       2      4       SED -> MLKA (IN MS11 / V18).
FT   VARIANT     234    234       V -> M (IN MS11 / V18).
FT   NON_TER     238    238
SQ   SEQUENCE   238 AA;  27073 MW;  883A3560C2DF1B9F CRC64;

  Query Match             8.2%;  Score 91.5;  DB 1;  Length 238;
  Best Local Similarity   24.7%;  Pred. No. 4.7;
  Matches   45;  Conservative   35;  Mismatches   69;  Indels   33;  Gaps    9;

Qy          4 SHHHHHHGSDDKVATITSGNKSTNVTVHKSEAGTSSVF------YYKTGDMLPEDTTHVR 57
              :: : |   |     |  : || : |: :      | ||       |   |  :  |    |
Db         16 AYAYEHITHDYPEPTAPNKNKISTVSDYFRNIRTRSVHPRVSVGYDFGGWRIAADYARYR 75

Qy         58 -WFLNINNEKSYVSKDITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFE---- 112
               |    || |  |: :  :: : : | ::|  | |      |:    | ||| ||:
Db         76 KW----NNNKYSVNIE-NVRIRKENGIRIDRKTENQENGTFHAVSSLGLSAIYDFQINDK 130

Qy        113 -KAFPGSKI-------TVDNTKNTIDV-TIPQG--YGSYNSFSINYKTKITNEQQKEFVN 161
               | : |:::       ::|:|| ||:| |:|     |:  ::: : ||      | :: :
Db        131 FKPYIGARVAYGHVRHSIDSTKKTIEVTTVPSNAPNGAVTTYNTDPKT------QNDYQS 184

Qy        162 NS 163
              ||
Db        185 NS 186


RESULT 12
GUN_BACSP
ID   GUN_BACSP      STANDARD;      PRT;   463 AA.
AC   P29019;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE   (Cellulase) (Endo-K).
OS   Bacillus sp. (strain KSM-330).
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1409;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92121880; PubMed=1770347;
RA   Ozaki K., Sumitomo N., Ito S.;
RT   "Molecular cloning and nucleotide sequence of the gene encoding an
RT   endo-1,4-beta-glucanase from Bacillus sp. KSM-330.";
RL   J. Gen. Microbiol. 137:2299-2305(1991).
RN   [2]
RP   SEQUENCE OF 56-75, AND CHARACTERIZATION.
RX   MEDLINE=91259037; PubMed=2045781;
RA   Ozaki K., Ito S.;
RT   "Purification and properties of an acid endo-1,4-beta-glucanase from
RT   Bacillus sp. KSM-330.";
RL   J. Gen. Microbiol. 137:41-48(1991).
CC   -!- FUNCTION: THIS ACID ENDOGLUCANASE IS ACTIVE OVER AN EXTREMELY
CC       NARROW RANGE OF PH VALUES, BETWEEN 4.5 AND 6.5, WITH AN OPTIMUM
CC       PH AT 5.2.
CC   -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC       linkages in cellulose, lichenin and cereal beta-D-glucans.
CC   -!- PTM: THE N- AND THE C-TERMINUS MAY BE SUBJECTED TO PROTEOLYSIS.
CC   -!- MISCELLANEOUS: ONE TRP RESIDUE HAS BEEN PROVED TO BE INVOLVED IN
CC       THE MECHANISM OF ACTION OF ENDO-K.
CC   -!- SIMILARITY: Belongs to cellulase family D (family 8 of glycosyl
CC       hydrolases).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M68872; AAA22409.1; -.
DR   PIR; A44808; A44808.
DR   InterPro; IPR002037; Glyco_hydro_8.
DR   InterPro; IPR008928; Glyco_trans_6hp.
DR   Pfam; PF01270; Glyco_hydro_8; 1.
DR   PRINTS; PR00735; GLHYDRLASE8.
DR   PROSITE; PS00812; GLYCOSYL_HYDROL_F8; 1.
KW   Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT   SIGNAL        1     27       POTENTIAL.
FT   PROPEP       28     55       POTENTIAL.
FT   CHAIN        56    463       ENDOGLUCANASE.
FT   ACT_SITE    130    130       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE    191    191       NUCLEOPHILE (POTENTIAL).
SQ   SEQUENCE   463 AA;  51882 MW;  407FA54F5236C59E CRC64;

  Query Match             8.1%;  Score 91;  DB 1;  Length 463;
  Best Local Similarity   21.2%;  Pred. No. 12;
  Matches   53;  Conservative   31;  Mismatches   64;  Indels  102;  Gaps   13;

Qy          7 HHHHGSDDKV-------ATITSGNKSTNVT----VHKSEAGTSSVFYYKTGDMLPED--T 53
              |   ||  |:          || | |::|||    ::  : |  | |     |  | |
Db        204 HKQWGSSGKINYLKEAQNMITKGIKASNVTKNNGLNLGDWGDKSTF-----DTRPSDWMM 258

Qy         54 THVRWFLNINNEKSYVSKDITIKDQIQGGQQLDLSTLNI--NVTGTHSNY---YSGQSA- 107
              :|:| |     :|::                        ||:  |:  |::|:   || ::
Db        259 SHLRAFYEFTGDKTW--------------------LNVIDNLYNTYTNFTNKYSPKTGL 297

Qy        108 ITDF---EKAFPGSKITVDNTKNT---------IDVTIPQGYGSYN-----------SFS 144
              |:||       |  |  :| :| |         : : |   |  |            :
Db        298 ISDFVVKNPPQPAPKDFLDESKYTDSYYYNASRVPLRIVMDYAMYGEKRGKVISDKVATW 357

Qy        145 INYKTK--------------------------------ITNEQQKEFVNNSQAWYQEHG 171
              |  |||                                || : :|:||:   |  ::
Db        358 IKSKTKGNPSKIVDGYKLDGTNIGDYPTAVYVSPFIAAGTTNSKNQEWVNSGWDW-MKNK 416

Qy        172 KEEVNGKSFN 181
              ||     |:|
Db        417 KESYFSDSYN 426


RESULT 13
LIPA_MYCPU
ID   LIPA_MYCPU     STANDARD;      PRT;   578 AA.
AC   Q50274; Q98Q39;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Lipoprotein A precursor.
GN   LIPA OR MYPU_5300.
OS   Mycoplasma pulmonis.
OC   Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2107;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KD735-15;
RX   MEDLINE=20245550; PubMed=10781561;
RA   Shen X., Gumulak J., Yu H., French C.T., Zou N., Dybvig K.;
RT   "Gene rearrangements in the vsa locus of Mycoplasma pulmonis.";
RL   J. Bacteriol. 182:2900-2908(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=UAB CTIP;
RX   MEDLINE=21267165; PubMed=11353084;
RA   Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
RA   Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
```

```
RA   Blanchard A.;
RT   "The complete genome sequence of the murine respiratory pathogen
RT   Mycoplasma pulmonis.";
RL   Nucleic Acids Res. 29:2145-2153(2001).
RN   [3]
RP   PARTIAL SEQUENCE FROM N.A.
RC   STRAIN=KD735-15;
RX   MEDLINE=96414471; PubMed=8817492;
RA   Bhugra B., Voelker L.L., Zou N., Yu H., Dybvig K.;
RT   "Mechanism of antigenic variation in Mycoplasma pulmonis: interwoven,
RT   site-specific DNA inversions.";
RL   Mol. Microbiol. 18:703-714(1995).
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC       (Potential).
CC   -!- SIMILARITY: BELONGS TO THE PULMONIS LIPAB LIPOPROTEIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U23947; AAB41030.2; -.
DR   EMBL; AL445565; CAC13703.1; -.
DR   PIR; B90578; B90578.
DR   MypuList; MYPU_5300; -.
DR   InterPro; IPR007326; Lipoprotein_17.
DR   InterPro; IPR000437; Prok_lipoprot_S.
DR   Pfam; PF04200; Lipoprotein_17; 1.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW   Lipoprotein; Membrane; Signal; Complete proteome; Palmitate.
FT   SIGNAL        1     27       POTENTIAL.
FT   CHAIN        28    578       Lipoprotein A.
FT   LIPID        28     28       N-palmitoyl cysteine (Potential).
FT   LIPID        28     28       S-diacylglycerol cysteine (Potential).
FT   DOMAIN       82    165       ASN-RICH.
FT   VARIANT       7      7       K -> R (IN STRAIN KD735-15).
FT   VARIANT      41     41       N -> S (IN STRAIN KD735-15).
FT   VARIANT     110    110       N -> D (IN STRAIN KD735-15).
FT   VARIANT     127    127       S -> A (IN STRAIN KD735-15).
FT   VARIANT     547    547       Y -> N (IN STRAIN KD735-15).
SQ   SEQUENCE   578 AA;  66219 MW;  B533ED3467005D89 CRC64;

  Query Match             8.1%;  Score 91;  DB 1;  Length 578;
  Best Local Similarity   20.2%;  Pred. No. 16;
  Matches   49;  Conservative   33;  Mismatches  103;  Indels   58;  Gaps    6;

Qy         14 DKVATITSGNK-STNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSKD 72
              ||   ::  || |||  |    : | |    ::     :|::     :   | |
Db         61 DKDKNVSQDNKDSTNKAVSNENSQTQSQKTNESSQNTKDDSSKTSNLITNQNSSSNTKSK 120

Qy         73 ITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVT 132
              |    |     | | :|  :||:         :    |   :   || |    ||   |
Db        121 IQENKQ----SQKDQNTSAVNVSALE------KQTKNDENISLVNSKDTNVILKNDEKVA 170

Qy        133 IPQGYG---SYNSFSINYKTKITNEQQKEFVNNSQ---AWYQE---------------- 169
              : :      | || ::| || : | | |  | : :    |:|:
Db        171 LAKDDSKEKSKNSSNLNLKTPVENRQNKNEVKDDKKALQWWQKLNESASILESFSYDQTS 230

Qy        170 ------------------------HGKEEVNGKSFNHTVHNINANAGIEGTVKGELKVLK 205
                                       | ::|:: |: |  | |:   :    : | ::|
Db        231 LSLTFKEGMPLGLEVVLKLENLDSHEEKEISFKTTNGKVQNVLLTSSNLTSGKWKIKSFS 290

Qy        206 QDK 208
              ||
Db        291 FDK 293


RESULT 14
C1GB_BACTZ
ID   C1GB_BACTZ     STANDARD;      PRT;  1169 AA.
AC   Q9ZAZ6;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Pesticidial crystal protein cry1Gb (Insecticidal delta-endotoxin
DE   CryIG(b)) (Crystaline entomocidal protoxin) (133 kDa crystal protein).
GN   CRY1GB OR CRYIG(B) OR CRYH2.
OS   Bacillus thuringiensis (subsp. wuhanensis).
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=52024;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HD-525;
RX   MEDLINE=20153386; PubMed=10688690;
RA   Kuo W.-S., Lin J.-H., Tzeng C.-C., Kao S.-S., Chak K.-F.;
RT   "Cloning of two new cry genes from Bacillus thuringiensis subsp.
RT   wuhanensis strain.";
RL   Curr. Microbiol. 40:227-232(2000).
CC   -!- FUNCTION: PROMOTES COLLOIDOSMOTIC LYSIS BY BINDING TO THE MIDGUT
CC       EPITHELIAL CELLS OF LEPIDOPTERAN LARVAE. TOXIC TO PIERIS RAPAE.
CC   -!- DEVELOPMENTAL STAGE: The crystal protein is produced during
CC       sporulation and is accumulated both as an inclusion and as part of
CC       the spore coat.
CC   -!- MISCELLANEOUS: Toxic segment of the protein is located in the N-
CC       terminus.
CC   -!- SIMILARITY: Belongs to the delta endotoxin family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U70725; AAD10291.1; -.
DR   HSSP; P02965; 1CIY.
DR   InterPro; IPR001178; Endotoxin.
DR   InterPro; IPR005638; endotoxin_C.
DR   InterPro; IPR005639; endotoxin_N.
DR   InterPro; IPR008979; Gal_bind_like.
DR   Pfam; PF00555; endotoxin; 1.
DR   Pfam; PF03944; endotoxin_C; 1.
DR   Pfam; PF03945; endotoxin_N; 1.
KW   Toxin; Sporulation.
SQ   SEQUENCE   1169 AA;  132904 MW;  D1EFC1508A8B10BD CRC64;

  Query Match             8.1%;  Score 90.5;  DB 1;  Length 1169;
  Best Local Similarity   24.5%;  Pred. No. 41;
  Matches   36;  Conservative   17;  Mismatches   81;  Indels   13;  Gaps    5;

Qy         28 VTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSKDITIKDQIQGGQQLDL 87
              || :|   |   |  ::      || |    | |   |: | :  :|  |     :: :
Db       1021 VTAYKEGYGEGCVTIHEI-----EDNTDELKFSNCVEEEIYPNNTVTCNDYTATQEEYE- 1074

Qy         88 STLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIPQGYGSYNSFSINY 147
              |    |    | | |   |:  |:     |    || :    :||| |      |
Db       1075 GTYTSRNRGYDGAYESNSSVPADYASAYEEKAYTDGRRDNTCESN--RGYGDYTPLPAGY 1132

Qy        148 KTKITNEQQKEFVNNSQAWYQEHGKEE 174
               ||    : : |    : |  | |: |
Db       1133 VTK----ELEYFPETDKVWI-EIGETE 1154


RESULT 15
BSL1_TRIVA
ID   BSL1_TRIVA     STANDARD;      PRT;   625 AA.
AC   Q8MTI2;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Putative surface protein bspA-like (TvBspA-like-625).
GN   BSPAL1.
```

```
OS   Trichomonas vaginalis.
OC   Eukaryota; Parabasalidea; Trichomonadida; Trichomonadidae;
OC   Trichomonadinae; Trichomonas.
OX   NCBI_TaxID=5722;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22356757; PubMed=12467987;
RA   Hirt R.P., Harriman N., Kajava A.V., Embley T.M.;
RT   "A novel potential surface protein in Trichomonas vaginalis contains
RT   a leucine-rich repeat shared by micro-organisms from all three
RT   domains of life.";
RL   Mol. Biochem. Parasitol. 125:195-199(2002).
CC   -!- FUNCTION: May bind host tissue.
CC   -!- SUBCELLULAR LOCATION: Surface membrane (Probable).
CC   -!- SIMILARITY: Contains 13 leucine-rich (LRR) repeats.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AY101349; AAM51159.1; -.
DR   InterPro; IPR007093; LRR_Tp.
KW   Repeat; Transmembrane; Leucine-rich repeat.
FT   DOMAIN       1    548       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   549    571       POTENTIAL.
FT   DOMAIN     572    625       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      38     60       LRR 1.
FT   REPEAT      61     83       LRR 2.
FT   REPEAT      85    106       LRR 3.
FT   REPEAT     107    129       LRR 4.
FT   REPEAT     153    175       LRR 5.
FT   REPEAT     176    198       LRR 6.
FT   REPEAT     200    221       LRR 7.
FT   REPEAT     222    245       LRR 8.
FT   REPEAT     247    267       LRR 9.
FT   REPEAT     271    293       LRR 10.
FT   REPEAT     325    347       LRR 11.
FT   REPEAT     348    368       LRR 12.
FT   REPEAT     369    392       LRR 13.
FT   DOMAIN     444    525       ASN/PRO-RICH.
FT   CARBOHYD    15     15       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   227    227       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   625 AA;  67021 MW;  62EF37BDA8AD001A CRC64;

  Query Match             8.0%;  Score 90;  DB 1;  Length 625;
  Best Local Similarity   19.9%;  Pred. No. 20;
  Matches   46;  Conservative   42;  Mismatches   73;  Indels   70;  Gaps   12;

Qy         14 DKVATITSG-----NKSTNVTVHKSEA----------GTSSVFYYKTGDMLPEDTTHVR 57
              | | || :      :| |::|: ||             :|:    | |:    || |
Db        232 DSVTTIANSAFYECSKLTSITIGKSVTRIEGNAFSKCYSLTSITIKTTNDITSSITTDV- 290

Qy         58 WFLNIN-NEKSYVSKDIT------IKDQIQGGQQLDLSTLNINVTGTHSN------YYSG 104
               |||    |  | :  :|       ||::        : :  |:  : ||        :
Db        291 -FLNCPITELIYETTGLTFLTYEYFKDKV--------TLIKFNIPKSDSNSMIRLQEITS 341

Qy        105 QSAITDFEKAFPGSKITVDNTKNTIDVTIPQGY------------------GSYNSFSI 145
                 :| |      :|:|::|     ::|||: :                    ::   ||
Db        342 LPTLTHFTNL---NKVTIENIN---ELTIPESFIEGDNFEILITNNIKSIDPNAFKDCSI 395

Qy        146 NYKTKI-TNEQQKEFVNNSQAWYQ-----EHGKEEVNGKSFNHTVHNINAN 190
              |  | : |:: : :|: |::: :     ::   |: | :  |     | |
Db        396 NKFTYLGTDKLENDFLKNAKSCEEVITSTKYSDNEIGGMTITHKQSEENPN 446


Search completed: July 27, 2004, 10:08:43
Job time : 15 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 27, 2004, 10:05:01 ; Search time 39 Seconds
(without alignments)
1707.034 Million cell updates/sec

Title: US-09-813-820-4
Perfect score: 1122
Sequence: 1 MRGSHHHHHHGSDDKVATIT.....GIEGTVKGELKVLKQDKDTK 211

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1035	92.2	1183	16	Q8NUH0	Q8nuh0 staphylococ
2	491	43.8	721	2	Q84BK5	Q84bk5 enterococcu
3	463.5	41.3	657	2	Q840A5	Q840a5 streptococc
4	462.5	41.2	221	2	Q8KRZ5	Q8krz5 streptococc
5	182.5	16.3	1151	2	Q7WWP3	Q7wwp3 arcanobacte
6	182	16.2	319	2	Q9F859	Q9f859 enterococcu
7	181	16.1	721	2	Q9F866	Q9f866 enterococcu
8	179	16.0	627	2	Q9F863	Q9f863 enterococcu
9	179	16.0	721	2	Q9F867	Q9f867 enterococcu
10	178	15.9	319	2	Q9EU70	Q9eu70 enterococcu
11	178	15.9	319	2	Q9F857	Q9f857 enterococcu
12	178	15.9	319	2	Q9F856	Q9f856 enterococcu
13	178	15.9	458	2	Q9XBQ7	Q9xbq7 enterococcu
14	178	15.9	580	2	Q9F862	Q9f862 enterococcu
15	178	15.9	580	2	Q9F861	Q9f861 enterococcu
16	178	15.9	627	2	Q9F864	Q9f864 enterococcu
17	178	15.9	674	2	Q9F865	Q9f865 enterococcu
18	178	15.9	674	16	Q836L2	Q836l2 enterococcu
19	177.5	15.8	319	2	Q9F858	Q9f858 enterococcu
20	176.5	15.7	787	2	Q83VG5	Q83vg5 erysipeloth
21	175	15.6	296	2	Q9F860	Q9f860 enterococcu
22	175	15.6	319	2	Q9EU80	Q9eu80 enterococcu
23	131	11.7	1092	2	Q54099	Q54099 streptococc
24	128	11.4	1991	2	Q83VG6	Q83vg6 erysipeloth
25	121	10.8	3381	5	Q8I2V4	Q8i2v4 plasmodium
26	119	10.6	336	16	Q9CHP4	Q9chp4 lactococcus
27	118	10.5	1361	5	Q8IBV5	Q8ibv5 plasmodium
28	116.5	10.4	1824	5	Q8I5A3	Q8i5a3 plasmodium
29	116	10.3	642	16	Q898I0	Q898i0 clostridium
30	114.5	10.2	1335	5	Q8I4Y7	Q8i4y7 plasmodium
31	114.5	10.2	1987	16	Q839R9	Q839r9 enterococcu
32	109.5	9.8	2770	5	Q8ILV0	Q8ilv0 plasmodium
33	108	9.6	723	16	Q8XP11	Q8xp11 clostridium
34	107.5	9.6	1231	16	Q88ZE8	Q88ze8 lactobacill
35	107.5	9.6	2577	5	Q8IAM1	Q8iam1 plasmodium
36	107	9.5	321	16	Q896Q5	Q896q5 clostridium
37	107	9.5	358	16	Q931E7	Q931e7 staphylococ
38	105.5	9.4	598	2	Q9JPR7	Q9jpr7 neisseria m
39	105	9.4	3135	5	Q8I3B5	Q8i3b5 plasmodium
40	104.5	9.3	1318	5	Q95PH4	Q95ph4 dictyosteli
41	104.5	9.3	2488	5	Q8I3R5	Q8i3r5 plasmodium
42	104	9.3	961	5	Q8IKT9	Q8ikt9 plasmodium
43	104	9.3	1605	5	Q8I4X5	Q8i4x5 plasmodium
44	104	9.3	1859	5	Q8IC27	Q8ic27 plasmodium
45	104	9.3	1936	5	Q8I5A6	Q8i5a6 plasmodium

ALIGNMENTS

RESULT 1
Q8NUH0
ID Q8NUH0 PRELIMINARY; PRT; 1183 AA.
AC Q8NUH0;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen adhesin.
GN MW2612.
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=196620;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.;
RT "Genome and virulence determinants of high virulence community-
RT acquired MRSA.";
RL Lancet 359:1819-1827(2002).
DR EMBL; AP004831; BAB96477.1; -.
DR PIR; A42404; A42404.
DR PIR; A48620; A48620.
DR InterPro; IPR008966; Adhes_bact.
DR InterPro; IPR008454; Cna_B.
DR InterPro; IPR008970; Cna_B_unit.
DR InterPro; IPR008456; Collagen_bind.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF05738; Cna_B; 7.
DR Pfam; PF05737; Collagen_bind; 1.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
KW Complete proteome.
SQ SEQUENCE 1183 AA; 133001 MW; 7A33DA876F9DA447 CRC64;

Query Match 92.2%; Score 1035; DB 16; Length 1183;
Best Local Similarity 98.5%; Pred. No. 7.6e-58;
Matches 197; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Qy          12 SDDKVATITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSK 71
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db         144 SDDKVATITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTTHVRWFLNINNEKRYVSK 203

Qy          72 DITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDV 131
               ||||||||||||||||||||||||||||||| :|||||||||||||||||||||||||||
Db         204 DITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGSNAITDFEKAFPGSKITVDNTKNTIDV 263

Qy         132 TIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINANA 191
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         264 TIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINANA 323

Qy         192 GIEGTVKGELKVLKQDKDTK 211
               ||||||||||||||||||||
Db         324 GIEGTVKGELKVLKQDKDTK 343


RESULT 2
Q84BK5
ID   Q84BK5      PRELIMINARY;      PRT;   721 AA.
AC   Q84BK5;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Collagen adhesin.
OS   Enterococcus faecium (Streptococcus faecium).
OC   Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX   NCBI_TaxID=1352;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TX2555;
RX   MEDLINE=22511562; PubMed=12622825;
RA   Nallapareddy S.R., Weinstock G.M., Murray B.E.;
RT   "Clinical isolates of Enterococcus faecium exhibit strain-specific
RT   collagen binding mediated by Acm, a new member of the MSCRAMM
RT   family.";
RL   Mol. Microbiol. 47:1733-1747(2003).
DR   EMBL; AY135217; AAN12397.1; -.
DR   InterPro; IPR008966; Adhes_bact.
DR   InterPro; IPR008970; Cna_B_unit.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   Pfam; PF00746; Gram_pos_anchor; 1.
DR   PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW   Collagen.
SQ   SEQUENCE   721 AA;  81545 MW;  DB5BB3B3DEF45673 CRC64;

  Query Match             43.8%;  Score 491;  DB 2;  Length 721;
  Best Local Similarity   48.5%;  Pred. No. 2e-23;
  Matches   95;  Conservative   35;  Mismatches   64;  Indels    2;  Gaps    2;

Qy          18 TITSGNKSTNVTVHKSEAG-TSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSKDITIK 76
               |:|||:|:  | | |  :| :|||||||||||||||| |:||||||||  :|| : : |
Db         150 TVTSGDKTATVNVTKPASGSSSSVFYYKTGDMLPEDTKHIRWFLNINNNGTYVEQPVKIS 209

Qy          77 DQIQGGQQLDLSTLNIN-VTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIPQ 135
               |:|| ||:|| ||  || :       | |:  |  | : || :      | | |::|||:
Db         210 DEIQSGQRLDPSTFEINQIHLGEQKVYRGEEGIQQFLQDFPSATFNFSVTDNYIEITIPK 269

Qy         136 GYGSYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINANAGIEG 195
                : :     ::||| | |  :|  | |:|:||::|  |  |:|:|||||| ||:|: |: |
Db         270 NFVNLRKIMVSYKTIIENPEQINFENHSEAWFKEFNKPAVDGESFNHTVKNISASGGVNG 329

Qy         196 TVKGELKVLKQDKDTK 211
               ||:||||: |   ||:
Db         330 TVRGELKIFKYINDTE 345


RESULT 3
Q840A5
ID   Q840A5      PRELIMINARY;      PRT;   657 AA.
AC   Q840A5;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Collagen-binding protein Cne precursor (Fragment).
GN   CNE.
OS   Streptococcus equi subsp. equi.
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=148942;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1866;
RA   Lannergard J., Frykberg L., Guss B.;
RT   "CNE, a collagen-binding protein of Streptococcus equi.";
RL   Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY193773; AAO43099.1; -.
DR   InterPro; IPR008966; Adhes_bact.
DR   InterPro; IPR008970; Cna_B_unit.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
KW   Collagen; Signal.
FT   SIGNAL       1     27       Potential.
FT   NON_TER    657    657
SQ   SEQUENCE   657 AA;  73889 MW;  A902E85BBCE59C4F CRC64;

  Query Match             41.3%;  Score 463.5;  DB 2;  Length 657;
  Best Local Similarity   47.7%;  Pred. No. 1e-21;
  Matches   94;  Conservative   36;  Mismatches   62;  Indels    5;  Gaps    3;

Qy          18 TITSGNKSTNVTVHKSEAGT-SSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSKDITIK 76
               |:  ||::  :|| | |||| :| |||||||| | ||  ||||| ||| | :|:  :|::
Db         146 TVKVGNRTATITVTKPEAGTGTSSFYYKTGDMQPNDTERVRWFLLINNNKEWVANTVTVE 205

Qy          77 DQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIPQG 136
               | ||||| ||:|: :| |:|  :  : |::|:|:|   || | ||   | | | | : |
Db         206 DDIQGGQTLDMSSFDITVSGYRNERFVGENALTEFHTTFPNSVITA--TDNHISVRLDQY 263

Qy         137 YGSYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINANAGIEGT 196
                 | |: :| ||||||:  |||| |||: |||   |::|:|:  || | ||||| |::|:
Db         264 DASQNTVNIAYKTKITDFDQKEFANNSKIWYQILYKDQVSGQESNHQVANINANGGVDGS 323

Qy         197 --VKGELKVLKQDKDTK 211
                 :| :  ||: :
Db         324 RYTSFTVKKIWNDKENQ 340


RESULT 4
Q8KRZ5
ID   Q8KRZ5      PRELIMINARY;      PRT;   221 AA.
AC   Q8KRZ5;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Putative collagen adhesin se73.9 (Fragment).
OS   Streptococcus equi.
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1336;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CF32;
RA   Qin A., Artiushin S.C., Timoney J.F.;
RT   "Identification and Genomic Organization of Genes for Immunoreactive
RT   Surface Exposed Proteins of Streptococcus equi.";
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF521599; AAM77657.1; -.
DR   InterPro; IPR008966; Adhes_bact.
DR   InterPro; IPR008970; Cna_B_unit.
DR   InterPro; IPR008456; Collagen_bind.
DR   Pfam; PF05737; Collagen_bind; 1.
KW   Collagen.
FT   NON_TER       1      1
FT   NON_TER     221    221
SQ   SEQUENCE   221 AA;  25049 MW;  0300F430D21EB859 CRC64;
```

```
  Query Match             41.2%;  Score 462.5;  DB 2;  Length 221;
  Best Local Similarity   47.7%;  Pred. No. 3.4e-22;
  Matches   94;  Conservative   36;  Mismatches   62;  Indels    5;  Gaps    3;

Qy         18 TITSGNKSTNVTVHKSEAGT-SSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSKDITIK 76
              |:  ||::  :|| | |||| :| |||||||| | ||  ||||| ||| | :|:  :|::
Db          2 TVKVGNRTCTITVTKPEAGTGTSSPYYKTGDMQPNDTERVRWFLLINNNKEWVANTVTVE 61

Qy         77 DQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIPQG 136
              | ||||| ||:|: :| |:|  :  : |::|:|:|   || | ||   | | | : |
Db         62 DDIQGGQTLDMSSFDITVSGYRNERFVGENALTEFHTTFPNSVITA--TDNHISVRLDQY 119

Qy        137 YGSYNSFSINYKTKITNEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINANAGIEGT 196
              | |: :| ||||||:  |||| |||: |||   |::|:|:  || | ||||| |::|:
Db        120 DASQNTVNIAYKTKITDFDQKEFANNSKIWYQILYKDQVSGQESNHQVANINANGGVDGS 179

Qy        197 --VKGELKVLKQDKDTK 211
                  :| :  ||: :
Db        180 RYTSPTVKKIWNDKENQ 196


RESULT 5
Q7WWP3
ID   Q7WWP3      PRELIMINARY;      PRT;  1151 AA.
AC   Q7WWP3;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Collagen-binding protein A precursor.
GN   CBPA.
OS   Arcanobacterium pyogenes.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Actinomycineae; Actinomycetaceae; Arcanobacterium.
OX   NCBI_TaxID=1661;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22756333; PubMed=12874314;
RA   Esmay P.A., Billington S.J., Link M.A., Songer J.G., Jost B.H.;
RT   "The Arcanobacterium pyogenes Collagen-Binding Protein, CbpA, Promotes
RT   Adhesion to Host Cells.";
RL   Infect. Immun. 71:4368-4374(2003).
DR   EMBL; AY223543; AAO43108.1; -.
KW   Collagen; Signal.
FT   SIGNAL       1     27       Potential.
SQ   SEQUENCE   1151 AA;  124719 MW;  3127352B9795CE98 CRC64;

  Query Match             16.3%;  Score 182.5;  DB 2;  Length 1151;
  Best Local Similarity   29.5%;  Pred. No. 0.0016;
  Matches   61;  Conservative   34;  Mismatches   91;  Indels   21;  Gaps    8;

Qy         18 TITSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTT-HVRWFLNIN-NEKSYVSKDITI 75
              |:  ||  | | :  :  |   ||  ||| : :    :|||||  | |   :   |: |
Db        156 TVYVGN--TAVQISHNRGGDRGVFAGKTGLIYGDSNPGYVRWFLRANINGDPWPGGDVDI 213

Qy         76 KDQIQGGQQLDLSTLNINVTGTHSN------YYSGQSAITDFEKAFPGSK----ITVDNT 125
              :| :  ||:|| : : |   | | |       |     :| ||  :  |::    |
Db        214 EDTLGIGQKLDGTGIRI---GLHWNGQRPGSYSLTWRSIEDFLNSDYGTRAGTTIAYKEH 270

Qy        126 KNTIDVTIPQGYGSYNSFSINYKTKITNEQQKEFVNNSQ-AWYQEHGKEEVNGKSFNHTV 184
              :  | : :||   :   ||  |  || :: |:|| | :   :|:   |: :      :||
Db        271 EGRISIHVPQSVVNGREFSFTYDAKIMDDTQEEFTNTATFDFYENQKKDRIVD---IYTV 327

Qy        185 HNINANAGIEGTVKGELKVLKQDKDTK 211
               |  |: ||||     | : |  ||::
Db        328 RNPKASGGIEGKTTASLNLDKVLKDSR 354


RESULT 6
Q9F859
ID   Q9F859      PRELIMINARY;      PRT;  319 AA.
AC   Q9F859;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Collagen adhesin (Fragment).
GN   ACE.
OS   Enterococcus faecalis (Streptococcus faecalis).
OC   Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX   NCBI_TaxID=1351;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Beirut/TX0645;
RX   MEDLINE=20407335; PubMed=10948146;
RA   Nallapareddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT   "Diversity of ace, a gene encoding a microbial surface component
RT   recognizing adhesive matrix molecules, from different strains of
RT   enterococcus faecalis and evidence for production of ace during human
RT   infections.";
RL   Infect. Immun. 68:5210-5217(2000).
DR   EMBL; AF260889; AAG23947.1; -.
DR   HSSP; Q53654; 1AMX.
DR   InterPro; IPR008966; Adhes_bact.
DR   InterPro; IPR008456; Collagen_bind.
DR   Pfam; PF05737; Collagen_bind; 1.
FT   NON_TER       1      1
FT   NON_TER     319    319
SQ   SEQUENCE   319 AA;  35986 MW;  D5F914E0856A6940 CRC64;

  Query Match             16.2%;  Score 182;  DB 2;  Length 319;
  Best Local Similarity   28.8%;  Pred. No. 0.00039;
  Matches   55;  Conservative   32;  Mismatches   88;  Indels   16;  Gaps    6;

Qy         18 TITSGNKSTNVTVHKSEAG---TSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSKDIT 74
              | |:  : |   |  :| |       |:|| ||:  | :  ||||||:|  || |::||:
Db        112 TATATQRLTIEGVTNTETGQIERDYPFFYKVGDLAGE-SNQVRWFLNVNLNKSDVTEDIS 170

Qy         75 IKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIP 134
              | |:   ||||:  :   ::       |    :: :||:   |    |  | |  ::
Db        171 IADRQGSGQQLNKESFTFDIVNDKETKY---ISLAEFEQQGYGKIDFV--TDNDFNLRFY 225

Qy        135 QGYGSYNSFSINYKTKIT--NEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINANAG 192
              |    : || : | : ||   :  |  | |:    || : ::  | |: :        |
Db        226 QDKARFTSFIVRYTSTITEAGQHQATFENSYDINYQLNNQDATNEKNTSQV-----KNVF 280

Qy        193 IEGTVKGELKV 203
              :||   |   |
Db        281 VEGEASGNQNV 291


RESULT 7
Q9F866
ID   Q9F866      PRELIMINARY;      PRT;  721 AA.
AC   Q9F866;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Collagen adhesin precursor.
GN   ACE.
OS   Enterococcus faecalis (Streptococcus faecalis).
OC   Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX   NCBI_TaxID=1351;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=END6/TX0045;
RX   MEDLINE=20407335; PubMed=10948146;
RA   Nallapareddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT   "Diversity of ace, a gene encoding a microbial surface component
RT   recognizing adhesive matrix molecules, from different strains of
RT   enterococcus faecalis and evidence for production of ace during human
RT   infections.";
RL   Infect. Immun. 68:5210-5217(2000).
DR   EMBL; AF260873; AAG23932.1; -.
DR   HSSP; Q53654; 1AMX.
DR   InterPro; IPR008966; Adhes_bact.
DR   InterPro; IPR008456; Collagen_bind.
DR   InterPro; IPR001899; Gram_pos_anchor.
```

```
DR   Pfam; PF05737; Collagen_bind; 1.
DR   Pfam; PF00746; Gram_pos_anchor; 1.
DR   PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW   Signal.
FT   SIGNAL        1     31       POTENTIAL.
SQ   SEQUENCE   721 AA;  79047 MW;  F09EDBE49E5DF866 CRC64;

  Query Match             16.1%;  Score 181;  DB 2;  Length 721;
  Best Local Similarity   28.8%;  Pred. No. 0.0012;
  Matches   55;  Conservative   32;  Mismatches   88;  Indels   16;  Gaps    6;

Qy         18 TITSGNKSTNVTVHKSEAG---TSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSKDIT 74
              | |:  : |   |  :| |       |:|| ||:  | :  ||||||:|  || |::||:
Db        152 TATATQRLTIEGVTNTETGQIERDYPFFYKVGDLAGE-SNQVRWFLNVNLNKSDVTEDIS 210

Qy         75 IKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIP 134
              | |:   ||||:  :   ::       |    :: :||:   |    |  | |  ::
Db        211 IADRQGSGQQLNKESFTFDIVNDKETKY---ISLAEFEQQGYGKIDFV--TDNDFNLRFY 265

Qy        135 QGYGSYNSFSINYKTKIT--NEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINANAG 192
              :   : || : | : ||   : |  | |:    || : ::  | |: :       |
Db        266 RDKARFTSFIVRYTSTITEAGQHQATFENSYDINYQLNNQEATNEKNTSQV-----KNVF 320

Qy        193 IEGTVKGELKV 203
              :||   |   |
Db        321 VEGEASGNQNV 331


RESULT 8
Q9F863
ID   Q9F863      PRELIMINARY;      PRT;   627 AA.
AC   Q9F863;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Collagen adhesin precursor.
GN   ACE.
OS   Enterococcus faecalis (Streptococcus faecalis).
OC   Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX   NCBI_TaxID=1351;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD9/TX0249;
RX   MEDLINE=20407335; PubMed=10948146;
RA   Nallapareddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT   "Diversity of ace, a gene encoding a microbial surface component
RT   recognizing adhesive matrix molecules, from different strains of
RT   enterococcus faecalis and evidence for production of ace during human
RT   infections.";
RL   Infect. Immun. 68:5210-5217(2000).
DR   EMBL; AF260877; AAG23935.1; -.
DR   HSSP; Q53654; 1AMX.
DR   InterPro; IPR008966; Adhes_bact.
DR   InterPro; IPR008456; Collagen_bind.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   Pfam; PF05737; Collagen_bind; 1.
DR   Pfam; PF00746; Gram_pos_anchor; 1.
DR   PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW   Signal.
FT   SIGNAL        1     31       POTENTIAL.
SQ   SEQUENCE   627 AA;  68848 MW;  78C5345DFD99F528 CRC64;

  Query Match             16.0%;  Score 179;  DB 2;  Length 627;
  Best Local Similarity   28.3%;  Pred. No. 0.0013;
  Matches   54;  Conservative   33;  Mismatches   88;  Indels   16;  Gaps    6;

Qy         18 TITSGNKSTNVTVHKSEAG---TSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSKDIT 74
              | |:  : |   |  :| |       |:|| ||:  | :  ||||||:|  || |::||:
Db        152 TATATQRLTIEGVTNTETGQIERDYPFFYKVGDLAGE-SNQVRWFLNVNLNKSDVTEDIS 210

Qy         75 IKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIP 134
              | |:   ||||:  :   ::       |    :: :||:   |    |  | |  ::
Db        211 IADRQGSGQQLNKESFTFDIVNDKETKY---ISLAEFEQQGYGKIDFV--TDNDFNLRFY 265

Qy        135 QGYGSYNSFSINYKTKIT--NEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINANAG 192
              :   : || : | : ||   : |  | |:    || : ::  | |: :       |
Db        266 RDKARFTSFIVRYTSTITEAGQHQATFENSYDINYQLNNQEATNEKNTSQV-----KNVF 320

Qy        193 IEGTVKGELKV 203
              :||   |   |
Db        321 VEGEASGNQNV 331


RESULT 9
Q9F867
ID   Q9F867      PRELIMINARY;      PRT;   721 AA.
AC   Q9F867;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Collagen adhesin precursor.
GN   ACE.
OS   Enterococcus faecalis (Streptococcus faecalis).
OC   Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX   NCBI_TaxID=1351;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OG1RF;
RX   MEDLINE=20407335; PubMed=10948146;
RA   Nallapareddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT   "Diversity of ace, a gene encoding a microbial surface component
RT   recognizing adhesive matrix molecules, from different strains of
RT   enterococcus faecalis and evidence for production of ace during human
RT   infections.";
RL   Infect. Immun. 68:5210-5217(2000).
DR   EMBL; AF260872; AAG23931.1; -.
DR   HSSP; Q53654; 1AMX.
DR   InterPro; IPR008966; Adhes_bact.
DR   InterPro; IPR008456; Collagen_bind.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   Pfam; PF05737; Collagen_bind; 1.
DR   Pfam; PF00746; Gram_pos_anchor; 1.
DR   PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW   Signal.
FT   SIGNAL        1     31       POTENTIAL.
SQ   SEQUENCE   721 AA;  78842 MW;  1CA8A1D98CE69392 CRC64;

  Query Match             16.0%;  Score 179;  DB 2;  Length 721;
  Best Local Similarity   28.3%;  Pred. No. 0.0015;
  Matches   54;  Conservative   33;  Mismatches   88;  Indels   16;  Gaps    6;

Qy         18 TITSGNKSTNVTVHKSEAG---TSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSKDIT 74
              | |:  : |   |  :| |       |:|| ||:  | :  ||||||:|  || |::||:
Db        152 TATATQRLTIEGVTNTETGQIERDYPFFYKVGDLAGE-SNQVRWFLNVNLNKSDVTEDIS 210

Qy         75 IKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIP 134
              | |:   ||||:  :   ::       |    :: :||:   |    |  | |  ::
Db        211 IADRQGSGQQLNKESFTFDIVNDKETKY---ISLAEFEQQGYGKIDFV--TDNDFNLRFY 265

Qy        135 QGYGSYNSFSINYKTKIT--NEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINANAG 192
              :   : || : | : ||   : |  | |:    || : ::  | |: :       |
Db        266 RNKARFTSFIVRYTSTITEAGQHQATFENSYDINYQLNNQDATNEKNTSQV-----KNVF 320

Qy        193 IEGTVKGELKV 203
              :||   |   |
Db        321 VEGEASGNQNV 331


RESULT 10
Q9EU70
ID   Q9EU70      PRELIMINARY;      PRT;   319 AA.
AC   Q9EU70;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Collagen adhesin precursor (Fragment).
```

```
GN   ACE.
OS   Enterococcus faecalis (Streptococcus faecalis).
OC   Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX   NCBI_TaxID=1351;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Various strains;
RX   MEDLINE=20407335; PubMed=10948146;
RA   Nallapareddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT   "Diversity of ace, a gene encoding a microbial surface component
RT   recognizing adhesive matrix molecules, from different strains of
RT   enterococcus faecalis and evidence for production of ace during human
RT   infections.";
RL   Infect. Immun. 68:5210-5217(2000).
DR   EMBL; AF260895; AAG23953.1; -.
DR   EMBL; AF260880; AAG23938.1; -.
DR   EMBL; AF260881; AAG23939.1; -.
DR   EMBL; AF260882; AAG23940.1; -.
DR   EMBL; AF260883; AAG23941.1; -.
DR   EMBL; AF260888; AAG23946.1; -.
DR   EMBL; AF260890; AAG23948.1; -.
DR   EMBL; AF260893; AAG23951.1; -.
DR   HSSP; Q53654; 1AMX.
DR   InterPro; IPR008966; Adhes_bact.
DR   InterPro; IPR008456; Collagen_bind.
DR   Pfam; PF05737; Collagen_bind; 1.
FT   NON_TER       1      1
FT   NON_TER     319    319
SQ   SEQUENCE   319 AA;  36031 MW;  2FCF56B11AC90D20 CRC64;

  Query Match             15.9%;  Score 178;  DB 2;  Length 319;
  Best Local Similarity   28.3%;  Pred. No. 0.0007;
  Matches   54;  Conservative   33;  Mismatches   88;  Indels   16;  Gaps    6;

Qy         18 TITSGNKSTNVTVHKSEAG---TSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSKDIT 74
              | |:  : |   |  :| |       |:|| ||:  | :  ||||||:|  || |::||:
Db        112 TATATQRLTIEGVTNTETGQIERDYPFFYKVGDLAGE-SNQVRWFLNVNLNKSDVTEDIS 170

Qy         75 IKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIP 134
              | |:   ||||:  :   ::       |    :: :||:   |   |  | |  ::
Db        171 IADRQGSGQQLNKESFTFDIVNDKETKY---ISLAEFEQQGYGKIDFV--TDNDFNLRFY 225

Qy        135 QGYGSYNSFSINYKTKIT--NEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINANAG 192
              :   : || : | : ||   : |  | |:    || : ::  | |: :        |
Db        226 RDKARFTSFIVRYTSTITEAGQHQATFENSYDINYQLNNQDATNEKNTSQV-----KNVF 280

Qy        193 IEGTVKGELKV 203
              :||   |   |
Db        281 VEGEASGNQNV 291


RESULT 11
Q9F857
ID   Q9F857      PRELIMINARY;      PRT;   319 AA.
AC   Q9F857;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Collagen adhesin (Fragment).
GN   ACE.
OS   Enterococcus faecalis (Streptococcus faecalis).
OC   Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX   NCBI_TaxID=1351;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Parker/TX2619;
RX   MEDLINE=20407335; PubMed=10948146;
RA   Nallapareddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT   "Diversity of ace, a gene encoding a microbial surface component
RT   recognizing adhesive matrix molecules, from different strains of
RT   enterococcus faecalis and evidence for production of ace during human
RT   infections.";
RL   Infect. Immun. 68:5210-5217(2000).
DR   EMBL; AF260894; AAG23952.1; -.
DR   HSSP; Q53654; 1AMX.
DR   InterPro; IPR008966; Adhes_bact.
DR   InterPro; IPR008456; Collagen_bind.
DR   Pfam; PF05737; Collagen_bind; 1.
FT   NON_TER       1      1
FT   NON_TER     319    319
SQ   SEQUENCE   319 AA;  36001 MW;  2054C2B11ECC5920 CRC64;

  Query Match             15.9%;  Score 178;  DB 2;  Length 319;
  Best Local Similarity   28.3%;  Pred. No. 0.0007;
  Matches   54;  Conservative   33;  Mismatches   88;  Indels   16;  Gaps    6;

Qy         18 TITSGNKSTNVTVHKSEAG---TSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSKDIT 74
              | |:  : |   |  :| |       |:|| ||:  | :  ||||||:|  || |::||:
Db        112 TATATQRLTIEGVTNTETGQIERDYPFFYKVGDLAGE-SNQVRWFLNVNLNKSDVTEDIS 170

Qy         75 IKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIP 134
              | |:   ||||:  :   ::       |    :: :||:   |   |  | |  ::
Db        171 IADRQGSGQQLNKESFTFDIVNDKETKY---ISLAEFEQQGYGKIDFV--TDNDFNLRFY 225

Qy        135 QGYGSYNSFSINYKTKIT--NEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINANAG 192
              :   : || : | : ||   : |  | |:    || : ::  | |: :        |
Db        226 RDKARFTSFIVRYTSTITEAGQHQATFENSYDINYQLNNQDATNEKNTSQV-----KNVF 280

Qy        193 IEGTVKGELKV 203
              :||   |   |
Db        281 VEGEASGNQNV 291


RESULT 12
Q9F856
ID   Q9F856      PRELIMINARY;      PRT;   319 AA.
AC   Q9F856;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Collagen adhesin (Fragment).
GN   ACE.
OS   Enterococcus faecalis (Streptococcus faecalis).
OC   Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX   NCBI_TaxID=1351;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B-343/TX2783;
RX   MEDLINE=20407335; PubMed=10948146;
RA   Nallapareddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT   "Diversity of ace, a gene encoding a microbial surface component
RT   recognizing adhesive matrix molecules, from different strains of
RT   enterococcus faecalis and evidence for production of ace during human
RT   infections.";
RL   Infect. Immun. 68:5210-5217(2000).
DR   EMBL; AF260896; AAG23954.1; -.
DR   HSSP; Q53654; 1AMX.
DR   InterPro; IPR008966; Adhes_bact.
DR   InterPro; IPR008456; Collagen_bind.
DR   Pfam; PF05737; Collagen_bind; 1.
FT   NON_TER       1      1
FT   NON_TER     319    319
SQ   SEQUENCE   319 AA;  36073 MW;  94562E9FDF88988A CRC64;

  Query Match             15.9%;  Score 178;  DB 2;  Length 319;
  Best Local Similarity   28.3%;  Pred. No. 0.0007;
  Matches   54;  Conservative   33;  Mismatches   88;  Indels   16;  Gaps    6;

Qy         18 TITSGNKSTNVTVHKSEAG---TSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSKDIT 74
              | |:  : |   |  :| |       |:|| ||:  | :  ||||||:|  || |::||:
Db        112 TATATQRLTIEGVTNTETGQIERDYPFFYKVGDLAGE-SNQVRWFLNVNLNKSDVTEDIS 170

Qy         75 IKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIP 134
              | |:   ||||:  :   ::       |    :: :||:   |   |  | |  ::
Db        171 IADRQGSGQQLNKESFTFDIVNDKETKY---ISLAEFEQQGYGKIDFV--TDNDFNLRFY 225
```

```
Qy        135 QGYGSYNSFSINYKTKIT--NEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINANAG 192
              :   : || : | : ||   : |  | |:    || : ::  | |: :       |
Db        226 RDKARFTSFIVRYTSTITEAGQHQATFENSYDINYQLNNQDATNEKNTSQV-----KNVF 280

Qy        193 IEGTVKGELKV 203
              :||   |   |
Db        281 VEGEASGNQNV 291
```

RESULT 13
Q9XBQ7
ID Q9XBQ7 PRELIMINARY; PRT; 458 AA.
AC Q9XBQ7;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen adhesin Ace (Fragment).
GN ACE.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CG110;
RX MEDLINE=99410431; PubMed=10480905;
RA Rich R.L., Kreikemeyer B., Owens R.T., LaBrenz S., Narayana S.V.,
RA Weinstock G.M., Murray B.E., Hook M.;
RT "Ace is a collagen-binding MSCRAMM from Enterococcus faecalis.";
RL J. Biol. Chem. 274:26939-26945(1999).
DR EMBL; AF159247; AAD43342.1; -.
DR HSSP; Q53654; 1AMX.
DR InterPro; IPR008966; Adhes_bact.
DR InterPro; IPR008456; Collagen_bind.
DR Pfam; PF05737; Collagen_bind; 1.
FT NON_TER 1 1
FT NON_TER 458 458
SQ SEQUENCE 458 AA; 51085 MW; B674692BD29A0A24 CRC64;

Query Match 15.9%; Score 178; DB 2; Length 458;
Best Local Similarity 28.3%; Pred. No. 0.0011;
Matches 54; Conservative 33; Mismatches 88; Indels 16; Gaps 6;

```
Qy         18 TITSGNKSTNVTVHKSEAG---TSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSKDIT 74
              | |:  : |   |  :| |       |:|| ||:  | :  ||||||:|  || |:::|:
Db        121 TATATQRLTIEGVTNTETGQIERDYPFFYKVGDLAGE-SNQVRWFLNVNLNKSDVTEDIS 179

Qy         75 IKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIP 134
              | |:   ||||:  :   ::       |    :: :||:   |    |  | |  ::
Db        180 IADRQGSGQQLNKESFTFDIVNDKETKY---ISLAEFEQQGYGKIDFV--TDNDFNLRFY 234

Qy        135 QGYGSYNSFSINYKTKIT--NEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINANAG 192
              :   : || : | : ||   : |  | |:    || : ::  | |: :       |
Db        235 RDKARFTSFIVRYTSTITEAGQHQATFENSYDINYQLNNQDATNEKNTSQV-----KNVF 289

Qy        193 IEGTVKGELKV 203
              :||   |   |
Db        290 VEGEASGNQNV 300
```

RESULT 14
Q9F862
ID Q9F862 PRELIMINARY; PRT; 580 AA.
AC Q9F862;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen adhesin precursor.
GN ACE.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LBJ-1/TX0020;
RX MEDLINE=20407335; PubMed=10948146;
RA Nallapareddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT "Diversity of ace, a gene encoding a microbial surface component
RT recognizing adhesive matrix molecules, from different strains of
RT enterococcus faecalis and evidence for production of ace during human
RT infections.";
RL Infect. Immun. 68:5210-5217(2000).
DR EMBL; AF260878; AAG23936.1; -.
DR HSSP; Q53654; 1AMX.
DR InterPro; IPR008966; Adhes_bact.
DR InterPro; IPR008456; Collagen_bind.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF05737; Collagen_bind; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW Signal.
FT SIGNAL 1 31 POTENTIAL.
SQ SEQUENCE 580 AA; 63948 MW; 7CD61EAD3FDA0993 CRC64;

Query Match 15.9%; Score 178; DB 2; Length 580;
Best Local Similarity 28.3%; Pred. No. 0.0014;
Matches 54; Conservative 33; Mismatches 88; Indels 16; Gaps 6;

```
Qy         18 TITSGNKSTNVTVHKSEAG---TSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSKDIT 74
              | |:  : |   |  :| |       |:|| ||:  | :  ||||||:|  || |:::|:
Db        152 TATATQRLTIEGVTNTETGQIERDYPFFYKVGDLAGE-SNQVRWFLNVNLNKSDVTEDIS 210

Qy         75 IKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIP 134
              | |:   ||||:  :   ::       |    :: :||:   |    |  | |  ::
Db        211 IADRQGSGQQLNKESFTFDIVNDKETKY---ISLAEFEQQGYGKIDFV--TDNDFNLRFY 265

Qy        135 QGYGSYNSFSINYKTKIT--NEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINANAG 192
              :   : || : | : ||   : |  | |:    || : ::  | |: :       |
Db        266 RDKARFTSFIVRYTSTITEAGQHQATFENSYDINYQLNNQDATNEKNTSQV-----KNVF 320

Qy        193 IEGTVKGELKV 203
              :||   |   |
Db        321 VEGEASGNQNV 331
```

RESULT 15
Q9F861
ID Q9F861 PRELIMINARY; PRT; 580 AA.
AC Q9F861;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen adhesin precursor.
GN ACE.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JH2-2;
RX MEDLINE=20407335; PubMed=10948146;
RA Nallapareddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT "Diversity of ace, a gene encoding a microbial surface component
RT recognizing adhesive matrix molecules, from different strains of
RT enterococcus faecalis and evidence for production of ace during human
RT infections.";
RL Infect. Immun. 68:5210-5217(2000).
DR EMBL; AF260879; AAG23937.1; -.
DR HSSP; Q53654; 1AMX.
DR InterPro; IPR008966; Adhes_bact.
DR InterPro; IPR008456; Collagen_bind.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF05737; Collagen_bind; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW Signal.
FT SIGNAL 1 31 POTENTIAL.

SQ SEQUENCE 580 AA; 64176 MW; 060EFE50D61A5271 CRC64;

Query Match 15.9%; Score 178; DB 2; Length 580;
Best Local Similarity 28.3%; Pred. No. 0.0014;
Matches 54; Conservative 33; Mismatches 88; Indels 16; Gaps 6;

```
Qy          18 TITSGNKSTNVTVHKSEAG---TSSVFYYKTGDMLPEDTTHVRWFLNINNEKSYVSKDIT 74
               | |:  : |   |  :| |       |:|| ||:  | :  ||||||:|  || |::||:
Db         152 TATATQRLTIEGVTNTETGQIERDYPFFYKVGDLAGE-SNQVRWFLNVNLNKSDVTEDIS 210

Qy          75 IKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDFEKAFPGSKITVDNTKNTIDVTIP 134
               | |:   ||||:  :   ::       |    :: :||:   |    |  | |  ::
Db         211 IADRQGSGQQLNKESFTFDIVNDKETKY---ISLAEFEQQGYGKIDFV--TDNDFNLRFY 265

Qy         135 QGYGSYNSFSINYKTKIT--NEQQKEFVNNSQAWYQEHGKEEVNGKSFNHTVHNINANAG 192
               :    : || : | : ||   : |  | |:    || : ::  | |: :        |
Db         266 RDKARFTSFIVRYTSTITEAGQHQATFENSYDINYQLNNQDATNEKNTSQV-----KNVF 320

Qy         193 IEGTVKGELKV 203
               :||   |   |
Db         321 VEGEASGNQNV 331
```

Search completed: July 27, 2004, 10:09:33
Job time : 42 secs